In one embodiment, the nucleotide and polypeptide sequences of the present invention may be used to design selective CA inhibitors. Studies have also shown that the different alpha-CA have different inhibitor binding properties (Sly et al., (1995), supra), suggesting that it may be possible to provide compounds that inhibit a CA isozyme of interest, such as CA II, while not binding to or inhibiting related enzymes such as the polypeptide of SEQ ID No. 390. The nucleic acid and polypeptide sequences of the invention can be used in computer based drug design or for carrying out binding predictions with candidate CA inhibitors in view of the extensive structural information publicly available for CA enzymes. In preferred embodiments, the nucleic acid and polypeptide of the invention is used in drug screening assays, including both cell based and non cell based assays. In one embodiment, a nucleotide or polypeptide sequence of the invention is brought into contact with a candidate CA inhibitor (such as a CA II inhibitor), and binding of the candidate inhibitor to the polypeptide of the invention, or the activity of the polypeptide of the invention is detected. Activity of the polypeptide of the invention may be CA activity, or any other suitable activity possessed by the polypeptide of the invention which may be inhibited by binding of the candidate substance. In preferred embodiments, a panel of CA isozymes including the polypeptide of the invention are screened against the candidate substance, including the polypeptide of SEQ ID No 39 and one or more enzymes selected from the group consisting of CA I, CA III, CA IV, CA VI, a CARP including but not limited to CARP VII, CARP X, CARP XI. In preferred embodiments, a candidate CA inhibitor is screened against one or more non-catalytic CA related proteins to eliminate undesired inhibition of these enzymes which may be involved in other important physiological functions. Means to conduct such drug screening assays are well known in the art.

Increasing alpha-CA activity for the treatment of alpha-CA deficiency disease

The polypeptide of the invention may also be used as a source of CA activity, such as for the treatment of disease. The defects in carbonic anhydrases are the cause of several diseases, including osteopetrosis (abnormally dense bone) renal tubular acidosis, cerebral calcification and mental retardation. Also, a carbonic anhydrase-related protein is described as being linked to cone-rod retinal distrophy (Bellinghan et al., 1998, Biochem. Biophys. Res. Comm.: 253, 364-367).

In one aspect, the invention thus involves increasing CA activity by providing increased activity of the polypeptide of SEQ ID No. 390. Increased activity of the polypeptide of SEQ ID No 390 can be provided by any suitable means, as further describer herein. Activity may be provided for example by introducing to a host cell or patient a vector containing a nucleotide sequence of SEQ ID No 149, treating said cell with a compound capable of increasing the expression of the polypeptide of the invention and/or treating a cell or patient directly with a polypeptide of SEQ ID No 390. In preferred embodiments, the polypeptide of the invention comprises at least one amino acid substitution, deletion or insertion. In one aspect, such amino acid changes are preferably in the catalytic site; preferably said amino acid changes involve the substitution, deletion or insertion of a

His residue and preferably said amino acid changes increase the CO_2 hydration activity of the polypeptide of the invention.

Metal ion biosensors

In further aspects, metal ion biosensors can be designed based on the polypeptide of SEQ ID No 390. Determination of metal ion concentrations in complex media such as serum, cell cytoplasm as well as for example seawater are important analytical functions that require high degrees of sensitivity and selectivity.

Biosensors may be particularly useful in detecting metal ion fluxes in and between cells. Such biosensors may exploit metal-binding ability of the polypeptide of the invention, as described by Thompson et al., who have developed such biosensors based on the CA enzyme (CA II). Such biosensors are useful in the detection of metal ion flux for example in the central nervous system. Zinc-containing neurons found throughout the mammalian cerebral cortex, striatum and amygdalar nuclei have been shown to release their zinc in a depolarization- and calcium-dependent fashion in vitro and in vivo. This zinc release has been suggested to act as a trans-synaptic neuromodulator : which has in turn been linked to excitotoxic neuronal cell death. CA based biosensors developed by Thomspon et al. showed that zinc is present and can be detected in extracellular medium from neurons. (Thompson et al, J. Neurosci Methods 96:35-45 (2000)).

Biosensors based on CA have been shown to be extremely selective, detecting Cu at subpicomolar levels, which is of sensitivity that might be achieved with mass spectometric techniques. Sensors based on the CA II isozyme have been shown to detect Zn and Cu at picomolar levels, and Cd, Co and Ni at nanomolar levels. (Thompson et al., Anal. Biochem. 267:185-195 (1999)). CA based biosensors have also demonstrated selectivity over potential interferents in biological systems at mM levels in extracellular fluids, such as Mg and Ca. (Thompson et al. (2000), supra).

Biosensors based on the polypeptide of the invention are based on the high selectivity and sensitivity of CA isozymes for zinc. Because the binding of Zn in the active site of the enzyme affects the enzyme's ability to bind a CA inhibitor, it is possible to use a CA inhibitor that exhibits a detectable change upon binding to the polypeptide of the invention to detect the fraction of polypeptide bound to the inhibitor, and therefore bound to Zn. The fraction of polypeptide with bound Zn in turn is determined by the concentrations of free Zn and the polypeptide of the invention, and the dissociation constant for zinc.

In one example, binding of the CA inhibitor to the polypeptide of the invention is detected by using a fluorescent inhibitor, whereby the inhibitor shows a detectable change in fluorescence emission wavelength of polarization upon binding to the polypeptide of the invention. In one example, a fluorescent sulfonamide is used, such as the fluorophore ABD-N (Thompson et al. (2000), supra).

Engineered CA enzymes

CA isozymes have been shown to have differing levels of catalytic activity and efficiency. In preferred embodiments, particularly for treatments which involve providing the increased activity of the polypeptide of SEQ ID No 390 or for use in metal ion biosensors, the polypeptide of the invention may be modified for increased CO_2 hydration and/or zinc binding.

In particular, studies have been carried out characterizing residues important for maximal CA activity, allowing CA isozymes to be designed having desired levels of activity. Important structural elements in CA isozymes for zinc binding, CO_2 hydration activity and stability are reviewed in Lindskog, Pharmacol. Ther. 74(1):1-20 (1997) and Sly (1995), supra. In one example, studies of CA III showed that changing the Phe198 residue to a Leu198 residue (as in CAII) resulted in a 25 fold increase in activity. (Chen et al., (1993), supra). Catalysis has also been greatly increased in CA II by replacing the Thr200 residue with His, as is normally found in CA I enzymes. Most dramatically, a CA-related protein (CA-RP) which in its native form was missing important residues at the catalytic site and had no detectable CO2 hydration activity at all was rendered an active CA by only two point mutations. (Sjoblom et al., FEBS Lett. 398: 322-325(1996)).

Thus, in embodiments where the polypeptide of the invention is used to provide a source of CO_2 hydration or for its zinc binding properties, it is advantageous to modify the polypeptide of the invention by introducing at least one amino acid substitution, deletion or insertion. In one aspect, such amino acid changes are preferably in the catalytic site; preferably said amino acid changes involve the substitution, deletion or insertion of a His residue and preferably said amino acid changes increase the CO2 hydration activity of the polypeptide of the invention. Optimal amino acid changes can be determined by the skilled artisan, particularly in view of sequence comparisons which can be carried out with the many well-characterized CA isozymes.

Protein of SEQ ID NO:252 (internal designation 105-089-3-0-G10-CS)

The protein of SEQ ID NO:252 is encoded by the cDNA of SEQ ID NO:11. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:252 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 105-089-3-0-G10-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:11 described throughout the present application also pertain to the human cDNA of clone 105-089-3-0-G10-CS. It is over represented in fetal brain.

The protein of SEQ ID NO:252 encoded by the cDNA of SEQ ID NO:11 is distributed primarily in the prostate and salivary gland. The protein of SEQ ID NO:252 is homologous to sequences described in PCT publication WO9827205-A2 (which describes a protein that was isolated from a human adult salivary gland cDNA library), PCT publication WO9839446-A2, PCT publication WO9839446-A2. The disclosures of each of the preceding PCT publications is incorporated herein by reference in their entireties.

The protein of SEQ ID NO:252 is also homologous to a polypeptide described in PCT publication WO9835229-A1, the disclosure of which is incorporated herein by reference in its entirety. Wo9835229-A1 describes a peptide of 27 amino acid residues that corresponds to 23/27 of a portion of the protein of SEQ ID NO:252 (amino acid 20-46). This corresponds to 85% identity with conserved changes (3 out of 4) yielding a 96% homology.

The protein described in WO 9835229 was identified in reflex tears that were collected from 12 non-contact lens wearing male and female humans. Reflex tears were stimulated by gently rubbing the nasal mucosa with a cotton wool tipped bud. Two different batches were collected from two different groups and examined by analytical and preparative 2-dimensional electrophoresis. After separation in the second dimension and transfer to PVDF membranes, identified protein spots (by 0.1% (w/v) Coomassie Blue) were loaded into a membrane-compatible Hewlett-Packard cartridge. Sequencing was conducted with a Model G1005A (Hewlett-Packard, CA) sequenator. One of the proteins identified migrated at 25 kDa and was revealed to have 5 isoforms of different pI. Two of these were N-terminally sequenced and gave the sequence of the above peptide with a pI of 5.0 and 4.4. The different isoforms indicate that this protein undergoes post-translational modifications, including sialylation or acylation. The presence of these isoforms in different degrees could reflect the disease status of the individual. Accordingly, one embodiment of the present invention relates to the detection or diagnosis of disease by determining the activity or level of the protein of SEQ ID NO:252 or a polynucleotide encoding the protein of SEQ ID NO:252 in an individual. For example, detection of the secreted protein of SEQ ID NO:252 in an individual may be accomplished non-invasively by measuring protein levels in bodily fluids into which the protein is secreted, such as tears and saliva. Such methods may be empolyed both in humans and in animals. It is probable that after the signal peptide is cleaved, the protein of SEQ ID NO:252 is secreted into bodily fluids including tears and probably saliva.

The protein of SEQ ID NO:252 can also be used for the screening of non-ocular diseases, by analyzing tears for marker proteins, particularly indicative of cancer and genetic disease. In addition, an altered chromatographic profile (e.g. 2D gel) of the isoforms of the protein of SEQ ID NO:252 may also indicate the disease state of an individual. For example, the levels of marker proteins in relation to the protein of SEQ ID NO:252 may be determined to evaluate whether the individual is suffering from a disease. Alternatively, tears may be analyzed for the levels of different isoforms of the protein of SEQ ID NO:252 to determine whether the pattern of such isoforms is indicative of disease.

The protein of SEQ ID NO:252 or fragments thereof may also be used as a lubricant or cleansing agent for the eyes. This protein can be included in contact lenses washing and storage solutions. This protein can also be useful as an ingredient in eye washing solutions (e.g. eye drops) used for everyday redness or healing after surgical/laser intervention. For example, the protein may be used to reduce eye inflammation. Alternatively, anti-bacterial properties may be exploited by

including the protein of SEQ ID NO:252 or fragments thereof in solutions, creams or ointments for the eyes, as well as creams or ointments in general for external applications.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:252, fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition in an individual. In such embodiments, the protein of SEQ ID NO:252, or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ ID NO:252. The protein of SEQ ID NO:252 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:252 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:252 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:252 or a cell or preparation containing the protein of SEQ ID NO:252 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:252 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:252 may be identified by contacting the protein of SEQ ID NO:252 or a cell or preparation containing the protein of SEQ ID NO:252 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify the source of a sample as, for example, saliva or tears, or to distinguish between two or more possible sources of a sample on the basis of the level of the protein of SEQ ID NO:252 in the sample. For example, the protein of SEQ ID NO:252 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such antibodies may then be used to identify tissues of unknown origin, for example, forensic samples,differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from saliva or tears or tissues other than saliva or tears to determine whether the test sample is from saliva or tears. Alternatively, the level of the protein of SEQ ID NO:252 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:252 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ

hybridization, Northern blots, dot blots or other technques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from saliva or tears or tissues other than saliva or tears to determine whether the test sample is from saliva or tears.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:252, including using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:252 or a fragment thereof may be fixed to a solid support, such as a chromatograpy matrix. A preparation containing cells expressing the protein of SEQ ID NO:252 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:252 or a fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:252. In such techniques, the level of the protein of SEQ ID NO:252 in an ill individual is measured using techniques such as those described herein. The level of the protein of 252 in the ill individual is compared to the level in normal individuals to determine whether the individual has a level of the protein of SEQ ID NO:252 which is indicative of disease.

Protein of SEQ ID NO:308 (internal designation 187-41-0-0-i21-CS)

The protein of SEQ ID NO:308 is encoded by the cDNA of SEQ ID NO:67. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:308 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 187-41-0-0-i21-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:67 described throughout the present application also pertain to the human cDNA of clone 187-41-0-0-i21-CS.

The protein of SEQ ID NO:308 is highly homologous to human secreted protein nf87_1 from PCT publication WO 9935252-A2 (the disclsoure of which is incorporated herein by reference in its entirety), to amino acids 26-129 of the human secreted protein SEQ ID NO:441 from PCT publication WO 9906548-A2 (the disclosure of which is incorporated herein by reference in its entirety), and to amino acids 26-114 of human secreted protein SEQ ID NO:439 from PCT publication WO 9906548-A2, the disclosure of which is incorporated herein by reference in its entirety. Thus, the protein of the invention appears to be a polymorphic variant of nf87_1. Since most of the proteins with high homology to the sequence of the invention have longer 5'termini, it is conceivable that the protein of the invention is a truncated/spliced variant of these proteins.

The protein of SEQ ID NO:308 was identified among the cDNAs from a library constructed from brain. Tissue distribution analysis through a BLAST analysis of databases shows that mRNA encoding this protein was found primarily in kidney, liver, and cancerous prostate.

The protein of SEQ ID NO:308 has chemical and structural homology to human interferon-inducible (IFI) protein isoforms p27 (63%), HIFI (50% identity), and to interferon-induced protein 6-16 precursor (IFI-6-16, 36%). Furthermore, the protein of the invention has structural homology (40% identity) to the human erythropoietin (EPO) primary response gene, EPRG3pt from PCT publication WO 9906063-A2, the disclosure of which is incorporated herein by reference in its entirety. Thus, the present invention relates to nucleic acid and amino acid sequences of a novel IFI protein and to the use of these sequences in the diagnosis, study, prevention and treatment of disease.

The protein of SEQ ID NO:308 comprises 105 amino acids. From the amino acid alignments and the hydrophobicity plots, it has a predicted signal peptide sequence spanning residues 31-43 and two predicted transmembrane domains spanning residues 17-37, and 48-68. Accordingly, one embodiment of the present invention is a polypeptide comprising the signal peptide and/or one or more of the transmembrane doamins.

Interferons (IFNs) are a part of the group of intercellular messenger proteins known as cytokines. α-IFN is the product of a multigene family of at least 16 members, whereas b-IFN is the product of a single gene. α- and β-IFNs are also known as type I IFNs. Type I IFNs are produced in a variety of cells types. Biosynthesis of type I IFNs is stimulated by viruses and other pathogens, and by various cytokines and growth factors. γ-IFN, also known as type II IFN, is produced in T-cells and natural killer cells. Antigens to which the organism has been sensitized stimulate biosynthesis of type II IFN. Both α- and γ-IFNs are immunomodulators and anti-inflammatory agents, activating macrophages, T-cells and natural killer cells.

IFNs are part of the body's natural defense to viruses and tumors. They exert these defenses by affecting the function of the immune system and by direct action on pathogens and tumor cells. IFNs mediate these multiple effects in part by inducing the synthesis of many cellular proteins. Some interferon-inducible (IFI) genes are induced equally well by α-, β- and γ-IFNs. Other IFI genes are preferentially induced by the type I or by the type II IFNs. The various proteins produced by IFI genes possess antitumor, antiviral and immunomodulatory functions. The expression of tumor antigens in cancer cells is increased by α-IFN, and renders the cancer cells more susceptible to immune rejection. The IFI proteins synthesized in response to viral infections are known to inhibit viral functions such as cell penetration, uncoating, RNA and protein synthesis, assembly and release (Hardman JG et al 25 (1996) The Pharmacological Basis of Therapeutics, McGraw-Hill, New York NY pp 1211-1214, the disclosure of which is incorporated herein by reference in its entirety). Type II IFN stimulates expression of major histocompatibility complex (MHC) proteins and is thus used in immune response enhancement.

The IFI gene known as 6-16 encodes an mRNA, which is highly induced by type I IFNs in a variety of human cells (Kelly JM et al (1986) EMBO J 5:1601-1606, the disclosure of which is incorporated herein by reference in its entirety). After induction, 6-16 mRNA constitutes as much as 0.1% of the total cellular mRNA. The 6-16 mRNA is present at only very low levels in the absence of type I IFN, and is only weakly induced by type II IFN. The 6-16 mRNA encodes a hydrophobic protein of 130 amino acids. The first 20 to 23 amino acids comprise a putative signal peptide. Protein 6-16 has at least two predicted transmembrane regions culminating in a negatively charged C-terminus.

The p27 gene encodes a protein with 41% amino acid sequence identity to the 6-16 protein. The p27 gene is expressed in some breast tumor cell lines and in a gastric cancer cell line. In other breast tumor cell lines, in the HeLa cervical cancer cell line, and in fetal lung fibroblasts, p27 expression occurs only upon α-IFN induction. In one breast tumor cell line, p27 is independently induced by estradiol and by IFN (Rasmussen UB et al (1993) Cancer Res 53:4096-4101, the disclosure of which is incorporated herein by reference in its entirety). Expression of p27 was analyzed in 21 primary invasive breast carcinomas, 1 breast cancer bone metastasis, and 3 breast fibroadenomas. High levels of p27 were found in about one-half of the primary carcinomas and in the bone metastasis, but not in the fibroadenomas. These observations suggest that certain breast tumors may produce high levels of, or have increased sensitivity to, type I IFN as compared to other breast tumors (Rasmussen UB et al, supra). In addition, the p27 gene expressed at significant levels in normal tissues including colon, stomach and lung, but not expressed in placenta, kidney, liver or skin. (Rasmussen UB et al, supra).

The small hydrophobic IFI gene products may contribute to viral resistance. A hepatitis-C virus (HCV)-induced gene, 130-51, was isolated from a cDNA library prepared from chimpanzee liver during the acute phase of the infection. The protein product of this gene has 97% identity to the human 6-16 protein (Kato T et al (1992) Virology 190:856-860, the disclosure of which is incorporated herein by reference in its entirety). The authors of the preceding paper suggest that HCV infection actively induces IFN expression, which in turn induces expression of IFI genes including 130-51. The IFI proteins synthesized in response to viral infections are known to inhibit viral functions such as penetration, uncoating, RNA or protein synthesis, assembly or release. The 130-51 protein may inhibit one or more of these functions in HCV. A particular virus may be inhibited in multiple functions by IFI proteins. In addition, the principle inhibitory effect exerted by IFI proteins differs among the virus families (Hardman JG, supra, p 1211, the disclosure of which is incorporated herein by reference).

The HIFI protein (PCT publication WO 9812223-A2, the disclosure of which is incorporated herein by reference in its entirety) is a human sequence identified among cDNAs from a library constructed from human neonatal kidney. Northern blot analysis using LIFESEQ™ database (Incyte Pharmaceuticas, Palo Alto, CA) shows that HIFI mRNA was found only in

neonatal kidney. The HIFI protein consists of 104 amino acids and has 55%, 45%, and 46% amino acid sequence identity to p27, 6-16 and 130-51, respectively.

Based on the chemical and structural homology between the protein of SEQ ID NO:308 and the small hydrophobic IFI proteins from human and chimpanzee, it is believed that the protein of SEQ ID NO:308 is synthesized when interferons are produced in infections, inflammation, autoimmune diseases etc. Interferons are produced in response to various cytokines and growth factors, in viral infections, inflammation, autoimmune diseases, and cancers. Accordingly, the protein of SEQ ID NO:308 or fragments thereof may be used in diagnosis and treatment of diseases such as, but not limited to, autoimmune disorders such as rheumatoid arthritis, Graves disease, systemic lupus erythematosus, autoimmune hepatitis, Wegener's granulomatosis, sarcoidosis, polyarthritis, pemphigus, pemphigoid, erythema multiform, Sjogren's syndrome, inflammatory bowel disease, multiple sclerosis, myasthenia gravis keratitis, scleritis, Type I diabetes, insulin-dependent diabetes mellitus, Lupus Nephritis, and allergic encephalomyelitis; proliferative disorders including various forms of cancer such as leukemias, lymphomas (Hodgkins and non-Hodgkins), sarcomas, melanomas, adenomas, carcinomas of solid tissue, hypoxic tumors, squamous cell carcinomas of the mouth, throat, larynx, and lung, genitourinary cancers such as cervical and bladder cancer, hematopoietic cancers, head and neck cancers, and nervous system cancers, benign lesions such as papillomas, atherosclerosis, angiogenesis; viral infections, in particular HCV and HIV infections, as well as other pathogen-induced infections (e.g. leishmania).

The protein of SEQ ID NO:308 or fragments thereof may also be used to treat conditions associated with inflammation or immune impairment (e. g. reumathoid and osteo arthritis and AIDS).

Another embodiment of the present invention relates to the use of the protein of SEQ ID NO:308 or fragments thereof to treat and/or prevent the ill-effect of bacterial infection during pregnancy in mammals, such as spontaneous abortion and maternal death. In a preferred embodiment, the protein of the invention may be used to counteract the effects of the bacterial endotoxin lipopolysaccharide (LPS). The methods for using such compositions is described in Dziegielewska and Andersen, Biol. Neonate, 74:372-5 (1998), the disclosure of which is incorporated herein by reference in its entirety.

Furthermore, the protein of SEQ ID NO:308 or fragments thereof are useful as a reagent for analyzing the control of gene expression by interferons and other cytokines in both normal and diseased cells. The protein of the SEQ ID NO:308 or fragments thereof may be used to identify specific molecules with which it binds such as agonists, antagonists or inhibitors.

Another embodiment of the present invention relates to methods of using the protein of SEQ ID NO:308 or fragments thereof to identify and/or quantify cytokines of the interferon family as well as other cytokines such as IL10 and tumor antigens, which may interact with the protein of the invention.

The protein of SEQ ID NO:308 or fragments thereof may also be included in pharmaceutical preparations for treating cancer or prevention/treatment of other diseases associated with changes in expression of the protein of the invention. In another embodiment of the present invention, the protein of SEQ ID NO 308 or fragments thereof is used to inhibit and/or modulate the effect of cytokines and related molecule such as Il-2, TNF alpha, CTLA4, CD28, and others, by preventing the binding of the endogenous cytokine to their natural receptors, thereby blocking cell proliferation or inhibitory signals generated by the ligand-receptor binding event.

The protein of SEQ ID NO:308 or fragments thereof is useful to correct defects in in vivo models of disease such as autoimmune, inflammation and tumor models, by injecting the protein either intra peritoneally intravenously, subcutaneously or directly in the diseased tissue.

The DNA encoding the protein of SEQ ID NO:308 or fragments thereof is useful in diagnostic assays for conditions/diseases associated with expression of the protein of the invention. The diagnostic assay is useful to distinguish between absence, presence, and excess expression of the protein of the invention and to monitor regulation of levels of the protein of the invention during therapeutic intervention. The DNA may also be incorporated into effective eukaryotic expression vectors and directly targeted to a specific tissue, organ, or cell population for use in gene therapy to treat the above mentioned conditions, including tumors and/or to correct disease- or genetic-induced defects in any of the above mentioned proteins including the protein of the invention. The DNA may also be used to design antisense sequences and ribozymes, which can be administered to modify gene expression in tumor and pathogen-infected cells and to influence expression of cytokines and growth factors. In vivo delivery of genetic constructs into subjects can be developed to the point of targeting specific cell types, such as tumor where expression of the protein of the invention may be affected or is modulating the expression and/or activity of other proteins such as cytokines, growth factors, their receptors and/or tumor antigens. It is also useful to detect unknown upstream sequences (e. g. promoters and regulatory elements) by standard techniques and for research into the control of gene expression by interferons and other cytokines, as well as growth and transcription factors in normal and diseased cells. Hybridization probes are useful to detect DNA encoding the protein of the invention (or closely related molecules) in biological samples, and for mapping the naturally occurring genomic sequence to a particular chromosome/chromosome region. The DNA may be used to generate and/or treat in vivo animal models of disease, including susceptibility or resistance to infection, inflammation, tumors and autoimmune conditions, as well as tumor therapy, based on vaccine, knock-out and transgene technologies.

Antibodies against the protein of SEQ ID NO:308 or fragments thereof are useful for the diagnosis of conditions and diseases associated with its expression and to quantify the protein of the invention (e. g. in assays to monitor patients during therapeutic intervention). Antibodies specific for the protein may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments produced by a Fab expression library. Neutralizing antibodies are especially

preferred for diagnostics and therapeutics. Diagnostic assays for the protein of the invention include methods utilizing the antibody and a label to detect the protein of the invention in human body fluids or extracts of cells or tissues.

The protein of the invention and its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any variety of drug screening techniques including high throughput. Methods which may be used to quantitate the expression of the nucleotide or protein of the invention include, but are not limited to, polymerase chain reaction (PCR), RT-PCR, RNAse protection, Northern and western blotting, enzyme-linked immunosorbent asay (ELISA), radioimmunoassay (RIA), fluorescent activated cell sorting (FACS), immunoprecipitation, and chromatography.

Under conditions of significant blood loss, EPO therapy, or both, iron-restricted erythropoiesis is evident. However, intravenous or oral iron therapy has substantial drawbacks. Moreover, traditional biochemical markers of storage iron in patients with anemia of chronic disease are unhelpful in the assessment of iron status (Lawrence T et al (2000) Blood 96:823-833, the disclosure of which is incorporated herein by reference in its entirety). As the protein of SEQ ID NO:308 bears homology to the human erythropoietin (EPO) primary response gene, EPRG3pt, it may be used to promote red blood cell formation or to monitor the value of safer intravenous iron preparations in patients with blood loss anemia, particularly those undergoing EPO therapy.

The hydrophobic IFI protein of SEQ ID NO:308 or fragments thereof may be used to diagnose conditions associated with its induction. For example, the protein of SEQ ID NO:308 or fragments thereof may be useful in the diagnosis and treatment of tumors, viral infections, inflammation, or conditions associated with impaired immunity, anemia of chronic blood loss or chronic disease, hemochromatosis, and EPO therapy. Furthermore, this protein may be used for investigating the control of gene expression by IFNs and other cytokines, as well as hormones and growth factors, in normal and diseased cells.

The protein of SEQ ID NO:308 or fragments thereof is useful to correct defects in in vivo models of disease such as autoimmune, inflammation, anemia, iron-overload and tumor models, by injecting the protein either intra peritoneally intravenously, subcutaneously or directly in the diseased tissue.

In addition, the protein of SEQ ID NO:308 is structurally related to other proteins having homology and/or structural similarity with human p27 (Rasmussen, U.B., et al., 1993, Cancer Research 53:4096-4101, the disclosure of which is incorporated herein by reference). Accordingly, the protein of brain, fetal brain, kidney, fetal kidney, or colon may be used to regulate the proliferation of EPO-dependent cells or the growth and development of erythroid and other hematopoietic lineages.

The protein of SEQ ID NO:308 or fragments thereof, or polynucleotides encoding the protein of SEQ ID NO:308 or fragments thereof, may be used to treat or ameliorate anemia of

chronic disease and chronic renal failure, polycythemia, cancer, AIDS, drug- and phlebotomy-induced anemias, hemochromatosis, erythropoiesis mediated by EPO therapy, and other conditions associated with altered activity or levels of the protein of SEQ ID NO:308.

In another embodiment, the present invention relates to methods for identifying agonists and antagonists/inhibitors using the protein of SEQ ID NO:308 or fragments thereof, and treating conditions with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate levels or activity of the protein of SEQ ID NO:308. In still another embodiment of the invention relates to the use of the protein SEQ ID NO:308, fragments therof or the DNA encoding the protein of SEQ ID NO:308 or fragments thereof to monitor the value of iron therapy in patients undergoing EPO therapy, or experiencing blood loss, or both.

The DNA encoding the protein of SEQ ID NO:308 or fragments thereof is useful in diagnostic assays for conditions/diseases associated with abnormal expression of the protein of SEQ ID NO:308. The diagnostic assay is useful to distinguish between absence, presence, and excess expression of the protein of the invention and to monitor regulation of levels of the protein of the invention during therapeutic intervention. The DNA may also be incorporated into effective eukaryotic expression vectors and directly targeted to a specific tissue, organ, or cell population for use in gene therapy to treat the above mentioned conditions, including tumors and/or to correct disease- or genetic-induced defects in any of the above mentioned proteins including the protein of the invention. The DNA may also be used to design antisense sequences and ribozymes, which can be administered to modify gene expression in tumor and pathogen-infected cells and to influence expression of cytokines, hormones and growth factors. In vivo delivery of genetic constructs into subjects can be developed to the point of targeting specific cell types, such as tumor where expression of the protein of the invention may be affected or is modulating the expression and/or activity of other proteins such as cytokines, growth factors, their receptors and/or tumor antigens. It is also useful to detect unknown upstream sequences (e. g. promoters and regulatory elements) by standard techniques and for research into the control of gene expression by interferons and other cytokines, as well as growth and transcription factors in normal and diseased cells. Hybridization probes are useful to detect DNA encoding the protein of the invention (or closely related molecules) in biological samples, and for mapping the naturally occurring genomic sequence to a particular chromosome/chromosome region. The DNA may be used to generate and/or treat in vivo animal models of disease, including susceptibility or resistance to infection, tumors, autoimmune conditions, anemia and iron-overload, as well as tumor therapy, based on vaccine, knock-out and transgene technologies.

Antibodies against the protein of SEQ ID NO:308 are useful for the diagnosis of conditions and disease associated with its expression and to quantify the protein of the invention (e. g. in assays to monitor patients during therapeutic intervention). Antibodies specific for the protein may

include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments produced by a Fab expression library. Neutralizing antibodies are especially preferred for diagnostics and therapeutics. Diagnostic assays for the protein of SEQ ID NO:308 include methods utilizing the antibody and a label to detect the protein of the invention in human body fluids or extracts of cells or tissues.

The protein of SEQ ID NO:308 and its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any variety of drug screening techniques including high throughput. Methods which may be used to quantitate the expression of the nucleotide or protein of the invention include, but are not limited to, polymerase chain reaction (PCR), RT-PCR, RNAse protection, Northern blotting, enzyme-linked immunosorbent asay (ELISA), radioimmunoassay (RIA), fluorescent activated cell sorting (FACS), immunoprecipitation, and chromatography.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:308, fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition in an individual. For example, the condition may be cancer, including breast cancer, viral infection, bacterial infection, inflammation, autoimmune disorders, rheumatoid arthritis, Graves disease, systemic lupus erythematosus, autoimmune hepatitis, Wegener's granulomatosis, sarcoidosis, polyarthritis, pemphigus, pemphigoid, erythema multiform, Sjogren's syndrome, inflammatory bowel disease, multiple sclerosis, myasthenia gravis keratitis, scleritis, Type I diabetes, insulin-dependent diabetes mellitus, Lupus Nephritis, and allergic encephalomyelitis; proliferative disorders including various forms of cancer such as leukemias, lymphomas (Hodgkins and non-Hodgkins), sarcomas, melanomas, adenomas, carcinomas of solid tissue, hypoxic tumors, squamous cell carcinomas of the mouth, throat, larynx, and lung, genitourinary cancers such as cervical and bladder cancer, hematopoietic cancers, head and neck cancers, and nervous system cancers, benign lesions such as papillomas, atherosclerosis, angiogenesis; viral infections, in particular HCV and HIV infections, as well as other pathogen-induced infections (e. g. leishmania).

In such embodiments, the protein of SEQ ID NO:308 , or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ ID NO:308. The protein of SEQ ID NO:308 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:308 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:308 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:308 or a cell or preparation containing the protein of SEQ ID NO:308 with a test agent and assaying whether the

test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:308 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:308 may be identified by contacting the protein of SEQ ID NO:308 or a cell or preparation containing the protein of SEQ ID NO:308 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify the source of a sample as, for example, brain, kidney, liver, or cancerous prostate, or to distinguish between two or more possible sources of a sample on the basis of the level of the protein of SEQ ID NO:308 in the sample. For example, the protein of SEQ ID NO:308 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells frombrain, kidney, liver, or cancerous prostate or tissues other than brain, kidney, liver, or cancerous prostate to determine whether the test sample is from brain, kidney, liver, or cancerous prostate. Alternatively, the level of the protein of SEQ ID NO:308 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:308 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other technques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from brain, kidney, liver, or cancerous prostate or tissues other than brain, kidney, liver, or cancerous prostate to determine whether the test sample is from brain, kidney, liver, or cancerous prostate.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:308, including using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:308 or a fragment thereof may be fixed to a solid support, such as a chromatograpy matrix. A preparation containing cells expressing the protein of SEQ ID NO:308 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The

support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:308 or a fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:308. In such techniques, the level of the protein of SEQ ID NO:308 in an ill individual is measured using techniques such as those described herein. The level of the protein of SEQ ID NO:308 in the ill individual is compared to the level in normal individuals to determine whether the individual has a level of the protein of SEQ ID NO:308 which is associated with disease.

Protein of SEQ ID NOs:289 and 307 (internal designations 175-1-3-0-E5-CS.cor and 187-39-0-0-k12-CS)

The protein of SEQ ID NO:289 is encoded by the cDNA of SEQ ID NO:48. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:289 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 175-1-3-0-E5-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:48 described throughout the present application also pertain to the human cDNA of clone 175-1-3-0-E5-CS.

The protein of the invention consists of 130 amino acids. From the amino acid alignments and the hydrophobicity plots, it has a predicted signal peptide sequence spanning residues 8-20 and four predicted transmembrane domains spanning residues 2-24, 42-61, 70-90 and 99-119. Accordingly, some embodiments of the present invention relate to polypeptides comprising the signal peptide and/or one or more of the transmembrane domains.

The protein of SEQ ID NO:289 encoded by the cDNA of SEQ ID NO:48 is homologous to SEQ ID NO: 4199 from EP 1033401-A2 (the disclosure of which is incorporated herein by reference in its entirety), a human secreted protein. Another protein, SEQ ID NO:307, encoded by the cDNA of SEQ ID NO:66, is a polymorphic variant of the protein of SEQ ID NO:289, and shares all of the herein-described functions and uses.

The present invention relates to a novel protein identified among the cDNAs from a library constructed from salivary gland, and to the use of the nucleic acid and amino acid sequences disclosed herein in the study, diagnosis, prevention, and treatment of disease. Tissue distribution analysis predicted by BLAST on databases shows that mRNA encoding this protein was found primarily in brain and fetal brain, with lower amounts in kidney, fetal kidney and colon.

Interferons (IFNs) are a part of the group of intercellular messenger proteins known as cytokines. α-IFN is the product of a multigene family of at least 16 members, whereas b-IFN is the product of a single gene. α- and β-IFNs are also known as type I IFNs. Type I IFNs are produced in a variety of cells types. Biosynthesis of type I IFNs is stimulated by viruses and other pathogens,

and by various cytokines and growth factors. γ-IFN, also known as type II IFN, is produced in T-cells and natural killer cells. Antigens to which the organism has been sensitized stimulate biosynthesis of type II IFN. Both α- and γ-IFNs are immunomodulators and anti-inflammatory agents, activating macrophages, T-cells and natural killer cells.

IFNs are part of the body's natural defense to viruses and tumors. They exert these defenses by affecting the function of the immune system and by direct action on pathogens and tumor cells. IFNs mediate these multiple effects in part by inducing the synthesis of many cellular proteins. Some interferon-inducible (IFI) genes are induced equally well by α-, β- and γ-IFNs. Other IFI genes are preferentially induced by the type I or by the type II IFNs. The various proteins produced by IFI genes possess antitumor, antiviral and immunomodulatory functions. The expression of tumor antigens in cancer cells is increased by α-IFN, and renders the cancer cells more susceptible to immune rejection. The IFI proteins synthesized in response to viral infections are known to inhibit viral functions such as cell penetration, uncoating, RNA and protein synthesis, assembly and release (Hardman JG et al 25 (1996) The Pharmacological Basis of Therapeutics, McGraw-Hill, New York NY pp 1211-1214, the disclosure of which is incorporated herein by reference in its entirety). Type II IFN stimulates expression of major histocompatibility complex (MHC) proteins and is thus used in immune response enhancement.

The protein of SEQ ID NO:289 is a small hydrophobic protein having chemical and structural homology to human interferon-inducible (IFI) protein isoforms 6-16 (97% identity), HIFI (44%), and p27 (33%), as well as 130-51, the chimpanzee homolog of 6-16 - (97%). Thus, the protein of SEQ ID NO:289 and the nucleic acid encoding it are polymorphic variants of 6-16 or the gene encoding 6-16. The protein of SEQ ID NO:289, fragments thereof, or nucleic acids encoding the protein of SEQ ID NO:289 or fragments thereof may be used in the diagnosis, study, prevention and treatment of disease as described below.

The IFI gene known as 6-16 encodes an mRNA, which is highly induced by type I IFNs in a variety of human cells (Kelly JM et al (1986) EMBO J 5:1601-1606, the disclosure of which is incorporated herein by reference in its entirety). After induction, 6-16 mRNA constitutes as much as 0.1% of the total cellular mRNA. The 6-16 mRNA is present at only very low levels in the absence of type I IFN, and is only weakly induced by type II IFN. The 6-16 mRNA encodes a hydrophobic protein of 130 amino acids. The first 20 to 23 amino acids comprise a putative signal peptide. Protein 6-16 has at least two predicted transmembrane regions culminating in a negatively charged C-terminus.

The p27 gene encodes a protein with 41% amino acid sequence identity to the 6-16 protein. The p27 gene is expressed in some breast tumor cell lines and in a gastric cancer cell line. In other breast tumor cell lines, in the HeLa cervical cancer cell line, and in fetal lung fibroblasts, p27 expression occurs only upon α-IFN induction. In one breast tumor cell line, p27 is independently induced by estradiol and by IFN (Rasmussen UB et al (1993) Cancer Res 53:4096-4101, the

disclosure of which is incorporated herein by reference in its entirety). Expression of p27 was analyzed in 21 primary invasive breast carcinomas, 1 breast cancer bone metastasis, and 3 breast fibroadenomas. High levels of p27 were found in about one-half of the primary carcinomas and in the bone metastasis, but not in the fibroadenomas. These observations suggest that certain breast tumors may produce high levels of, or have increased sensitivity to, type I IFN as compared to other breast tumors (Rasmussen UB et al, supra). In addition, the p27 gene expressed at significant levels in normal tissues including colon, stomach and lung, but not expressed in placenta, kidney, liver or skin. (Rasmussen UB et al, supra).

The small hydrophobic IFI gene products may contribute to viral resistance. A hepatitis-C virus (HCV)-induced gene, 130-51, was isolated from a cDNA library prepared from chimpanzee liver during the acute phase of the infection. The protein product of this gene has 97% identity to the human 6-16 protein (Kato T et al (1992) Virology 190:856-860, the disclosure of which is incorporated herein by reference in its entirety). The authors of this paper suggest that HCV infection actively induces IFN expression, which in turn induces expression of IFI genes including 130-51. The IFI proteins synthesized in response to viral infections are known to inhibit viral functions such as penetration, uncoating, RNA or protein synthesis, assembly or release. The 130-51 protein may inhibit one or more of these functions in HCV. A particular virus may be inhibited in multiple functions by IFI proteins. In addition, the principle inhibitory effect exerted by IFI proteins differs among the virus families (Hardman JG, supra, p 1211, the disclosure of which is incorporated herein by reference).

The HIFI protein (PCT publication WO 9812223-A2, the disclosure of which is incorporated herein by reference in its entirety) is a human sequence identified among cDNAs from a library constructed from human neonatal kidney. Northern blot analysis using LIFESEQ™ database (Incyte Pharmaceuticas, Palo Alto, CA) shows that HIFI mRNA was found only in neonatal kidney. The HIFI protein consists of 104 amino acids and has 55%, 45%, and 46% amino acid sequence identity to p27, 6-16 and 130-51, respectively.

The hydrophobic IFI proteins of the invention may provide the basis for clinical diagnosis of diseases associated with their induction. These proteins may be useful in the diagnosis and treatment of tumors, viral infections, inflammation, or conditions associated with impaired immunity. Furthermore, these proteins may be used for investigations of the control of gene expression by IFNs and other cytokines in normal and diseased cells.

Based on the chemical and structural homology among the protein of SEQ ID NO:289 and the small hydrophobic IFI proteins from human and chimpanzee, it is believed that the protein of SEQ ID NO:289 is synthesized when interferons are produced in infections, inflammation, autoimmune diseases etc. Interferons are produced in response to various cytokines and growth factors, in viral infections, inflammation, autoimmune diseases, and cancers. Accordingly, the protein of SEQ ID NO:289 or fragments thereof may be used in diagnosis and treatment of diseases

such as, but not limited to, autoimmune disorders such as rheumatoid arthritis, Graves disease, systemic lupus erythematosus, autoimmune hepatitis, Wegener's granulomatosis, sarcoidosis, polyarthritis, pemphigus, pemphigoid, erythema multiform, Sjogren's syndrome, inflammatory bowel disease, multiple sclerosis, myasthenia gravis keratitis, scleritis, Type I diabetes, insulin-dependent diabetes mellitus, Lupus Nephritis, and allergic encephalomyelitis; proliferative disorders including various forms of cancer such as leukemias, lymphomas (Hodgkins and non-Hodgkins), sarcomas, melanomas, adenomas, carcinomas of solid tissue, hypoxic tumors, squamous cell carcinomas of the mouth, throat, larynx, and lung, genitourinary cancers such as cervical and bladder cancer, hematopoietic cancers, head and neck cancers, and nervous system cancers, benign lesions such as papillomas, atherosclerosis, angiogenesis; viral infections, in particular HCV and HIV infections, as well as other pathogen-induced infections (e. g. leishmania).

The protein of SEQ ID NO:289 or fragments thereof may also be used to treat conditions associated with inflammation or immune impairment (e. g. reumathoid and osteo arthritis and AIDS).

Another embodiment of the present invention relates to the use of the protein of SEQ ID NO:289 or fragments thereof to treat and/or prevent the ill-effect of bacterial infection during pregnancy in mammals, such as spontaneous abortion and maternal death. In a preferred embodiment, the protein of the invention may be used to counteract the effects of the bacterial endotoxin lipopolysaccharide (LPS). The methods for using such compositions is described in Dziegielewska and Andersen, Biol. Neonate, 74:372-5 (1998), the disclosure of which is incorporated herein by reference in its entirety.

Furthermore, the protein of SEQ ID NO:289 or fragments thereof are useful as a reagent for analyzing the control of gene expression by interferons and other cytokines in both normal and diseased cells. The protein of the SEQ ID NO:289 or fragments thereof may be used to identify specific molecules with which it binds such as agonists, antagonists or inhibitors.

Another embodiment of the present invention relates to methods of using the protein of SEQ ID NO:289 or fragments thereof to identify and/or quantify cytokines of the interferon family as well as other cytokines such as IL-10 and tumor antigens, which may interact with the protein of the invention.

The protein of SEQ ID NO:289 or fragments thereof may also be included in pharmaceutical preparations for treating cancer or prevention/treatment of other diseases associated with changes in expression of the protein of the invention. In another embodiment of the present invention, the protein of SEQ ID NO:289 or fragments thereof is used to inhibit and/or modulate the effect of cytokines and related molecule such as Il-2, TNF alpha, CTLA4, CD28, and others, by preventing the binding of the endogenous cytokine to their natural receptors, thereby blocking cell proliferation or inhibitory signals generated by the ligand-receptor binding event.

The protein of SEQ ID NO:289 or fragments thereof is useful to correct defects in in vivo models of disease such as autoimmune, inflammation and tumor models, by injecting the protein either intra peritoneally intravenously, subcutaneously or directly into the diseased tissue.

The DNA encoding the protein of SEQ ID NO:289 or fragments thereof is useful in diagnostic assays for conditions/diseases associated with expression of the protein of the invention. The diagnostic assay is useful to distinguish between absence, presence, and excess expression of the protein of the invention and to monitor regulation of levels of the protein of the invention during therapeutic intervention. The DNA may also be incorporated into effective eukaryotic expression vectors and directly targeted to a specific tissue, organ, or cell population for use in gene therapy to treat the above mentioned conditions, including tumors and/or to correct disease- or genetic-induced defects in any of the above mentioned proteins including the protein of the invention. The DNA may also be used to design antisense sequences and ribozymes, which can be administered to modify gene expression in tumor and pathogen-infected cells and to influence expression of cytokines and growth factors. In vivo delivery of genetic constructs into subjects can be developed to the point of targeting specific cell types, such as tumor where expression of the protein of the invention may be affected or is modulating the expression and/or activity of other proteins such as cytokines, growth factors, their receptors and/or tumor antigens. It is also useful to detect unknown upstream sequences (e. g. promoters and regulatory elements) by standard techniques and for research into the control of gene expression by interferons and other cytokines, as well as growth and transcription factors in normal and diseased cells. Hybridization probes are useful to detect DNA encoding the protein of the invention (or closely related molecules) in biological samples, and for mapping the naturally occurring genomic sequence to a particular chromosome/chromosome region. The DNA may be used to generate and/or treat in vivo animal models of disease, including susceptibility or resistance to infection, inflammation, tumors and autoimmune conditions, as well as tumor therapy, based on vaccine, knock-out and transgene technologies.

Antibodies against the protein of SEQ ID NO:289 or fragments thereof are useful for the diagnosis of conditions and diseases associated with its expression and to quantify the protein of the invention (e. g. in assays to monitor patients during therapeutic intervention). Antibodies specific for the protein may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments produced by a Fab expression library. Neutralizing antibodies are especially preferred for diagnostics and therapeutics. Diagnostic assays for the protein of the invention include methods utilizing the antibody and a label to detect the protein of the invention in human body fluids or extracts of cells or tissues.

The protein of the invention and its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any variety of drug screening techniques including high throughput. Methods which may be used to quantitate the expression of the nucleotide or protein of the invention include, but are not limited to, polymerase chain reaction

(PCR), RT-PCR, RNAse protection, Northern and western blotting, enzyme-linked immunosorbent asay (ELISA), radioimmunoassay (RIA), fluorescent activated cell sorting (FACS), immunoprecipitation, and chromatography.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:289, fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition in an individual. For example, the condition may be cancer, including breast cancer, viral infection, bacterial infection, inflammation, autoimmune disorders, rheumatoid arthritis, Graves disease, systemic lupus erythematosus, autoimmune hepatitis, Wegener's granulomatosis, sarcoidosis, polyarthritis, pemphigus, pemphigoid, erythema multiform, Sjogren's syndrome, inflammatory bowel disease, multiple sclerosis, myasthenia gravis keratitis, scleritis, Type I diabetes, insulin-dependent diabetes mellitus, Lupus Nephritis, and allergic encephalomyelitis; proliferative disorders including various forms of cancer such as leukemias, lymphomas (Hodgkins and non-Hodgkins), sarcomas, melanomas, adenomas, carcinomas of solid tissue, hypoxic tumors, squamous cell carcinomas of the mouth, throat, larynx, and lung, genitourinary cancers such as cervical and bladder cancer, hematopoietic cancers, head and neck cancers, and nervous system cancers, benign lesions such as papillomas, atherosclerosis, angiogenesis; viral infections, in particular HCV and HIV infections, as well as other pathogen-induced infections (e. g. leishmania).

In such embodiments, the protein of SEQ ID NO:289, or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ ID NO:289. The protein of SEQ ID NO:289 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:289 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:289 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:289 or a cell or preparation containing the protein of SEQ ID NO:289 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:289 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:289 may be identified by contacting the protein of SEQ ID NO:289 or a cell or preparation containing the protein of SEQ ID NO:289 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify the source of a sample as, for example, brain, fetal brain, kidney, fetal kidney, or colon, or to distinguish between two or more possible sources of a sample on the basis of the level of the protein of SEQ ID NO:289 in the sample. For example, the protein of SEQ ID NO:289 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from brain, fetal brain, kidney, fetal kidney, or colon or tissues other than brain, fetal brain, kidney, fetal kidney, or colon to determine whether the test sample is from brain, fetal brain, kidney, fetal kidney, or colon. Alternatively, the level of the protein of SEQ ID NO:289 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:289 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other technques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from brain, fetal brain, kidney, fetal kidney, or colon or tissues other than brain, fetal brain, kidney, fetal kidney, or colon to determine whether the test sample is from brain, fetal brain, kidney, fetal kidney, or colon.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:289, including using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:289 or a fragment thereof may be fixed to a solid support, such as a chromatograpy matrix. A prepartation containing cells expressing the protein of SEQ ID NO:289 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:289 or a fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:289. In such techniques, the level of the protein of SEQ ID NO:289 in an ill individual is measured using techniques such as those described herein. The level of the protein of SEQ ID NO:289 in the ill individual is compared to the level in normal individuals to determine whether the individual has a level of the protein of SEQ ID NO:289 which is associated with disease.

Protein of SEQ ID NO:268 (internal designation 116-111-4-0-B3-CS)

The protein of SEQ ID NO:268 is encoded by the cDNA of SEQ ID NO:27. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:268 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 116-111-4-0-B3-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:27 described throughout the present application also pertain to the human cDNA of clone 116-111-4-0-B3-CS. The protein of the invention is found to be expressed in testis and lungs.

The protein of SEQ ID NO:268 encoded by the extended cDNA SEQ ID NO: 27 is a splicing variant of XAGE-1, a member of the CT antigen family overexpressed in Ewing sarcoma (Liu, X. F., L. J. Helman, et al. (2000). Cancer Res 60(17): 4752-5, the disclosures of which are incorporated by reference herein in their entireties). In addition, the protein of SEQ ID NO:268 also shows strong homology at the COOH end with PAGE4, another member of the CT antigen family (Brinkmann, U., G. Vasmatzis, et al. (1999) Cancer Res 59(7): 1445-8, the disclosure of which is incorporated herein by reference in its entirety).

The cDNA SEQ ID NO:27 is composed of 5 exons. Exon 1 lies between nucleotides 1-245, exon2 lies between nucleotides 246-370, exon 3 lies between nucleotides 371-512, exon 4 lies between nucleotides 513-639, and exon 5 lies between nucleotides 640-762 . Exons 2 to 5 of cDNA SEQ ID NO:27 are shared in part with XAGE-1. However, since the initiation codon of SEQ ID NO: 27 is located in intron1 of XAGE-1, there is a frameshift in the alignment of the 2 molecules. Exon 1 of SEQ ID NO:27 lies between nucleotides 110-234 of XAGE-1, exon 2 of SEQ ID NO:27 lies between nucleotides 235-376 of XAGE-1, exon 3 of SEQ ID NO:27 lies between nucleotides 377-503 of XAGE-1, and exon 4 of SEQ ID NO:27 lies between nucleotides 504-526 of XAGE-1.

XAGE-1 is overexpressed in sarcoma and alveolar rhabdomyosarcoma and is also highly expressed in normal testis (Liu, X. F., L. J. Helman, et al. (2000). Cancer Res 60(17): 4752-5, the disclosure of which is incorporated herein by reference in its entirety). In addition XAGE-1 share homology with PAGE-4 (Brinkmann, U., G. Vasmatzis, et al. (1999) Cancer Res 59(7): 1445-8, the disclosure of which is incorporated herein by reference in its entirety) at the COOH end.

CT antigens are a distinct class of differenctiation antigens that are expressed by cancers arising in nonessential normal tissues such as prostate, breast, and ovary (G. Vasmatzis et al., Proc. Natl. Acad. Sci. USA, 95: 300-304, 1998, the disclosure of which is incorporated herein by reference in its entirety) and that have a restricted pattern of expression in normal tissues. This class of antigens are presented on the surface of tumor cells and are recognized by cytolytic T cells, leading to lysis. The extent to which these antigens have been studied, has been via cytolytic T cell characterization studies, in vitro i.e., the study of the identification of the antigen by a particular cytolytic T cell ("CTL" hereafter) subset. The subset proliferates upon recognition of the presented tumor rejection antigen, and the cells presenting the antigen are lysed. Characterization studies have

identified CTL clones which specifically lyse cells expressing the antigens. Examples of this work may be found in Levy et al., Adv. Cancer Res. 24: 1-59 (1977); Boon et al., J. Exp. Med. 152: 1184-1193 (1980); Brunner et al., J. Immunol. 124: 1627-1634 (1980) ; Maryanski et al., Eur. J. Immunol. 124: 1627-1634 (1980); Maryanski et al., Eur. J. Immunol. 12: 406-412 (1982); Palladino et al., Canc. Res. 47: 5074-5079 (1987), the disclosures of which are incorporated herein by reference in their entireties.

Some throughly studied CT antigens are MAGE, BAGE, GAGE and LAGE, others have been added including PAGE, XAGE, most of them located on chromosome X. Brinkmann et Al reported the identification of three new members of the GAGE/ PAGE family, termed XAGEs. XAGE-1 and XAGE-2 are expressed in Ewing's sarcoma, rhabdomyosarcoma, a breast cancer, and a germ cell tumor.

It is believed that the protein of SEQ ID NO:268 is a splicing variant of XAGE-1, a CT antigen overexpressed in Ewing sarcoma.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:268, fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition, such as those listed above, associated with over or under expression of the protein of SEQ ID NO:268. In such embodiments, the protein of SEQ ID NO:268, or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activity of the protein of SEQ ID NO:268. The protein of SEQ ID NO:268 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:268 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:268 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:268 or a cell or preparation containing the protein of SEQ ID NO:268 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:268 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:268 may be identified by contacting the protein of SEQ ID NO:268 or a cell or preparation containing the protein of SEQ ID NO:268 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably testis and lungs, or to distinguish between two or more possible sources of a tissue sample on the basis of the

level of the protein of SEQ ID NO:268 in the sample. For example, the protein of SEQ ID NO:268 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples,differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a tissue sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from testis or lungs or tissues other than testis or lungs to determine whether the test sample is from testis or lungs. Alternatively, the level of the protein of SEQ ID NO:268 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:268 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other technques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from testis or lungs or tissues other than testis or lungs to determine whether the test sample is from testis or lungs.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:268, including Ewing sarcoma cells, rhabdomyosarcoma cells, breast cancer cells and germ cell tumor cells using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:268 or a fragment thereof may be fixed to a solid support, such as a chromatograpy matrix. A prepartation containing cells expressing the protein of SEQ ID NO:268 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:268 or a fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:268. In some embodiments, the protein of SEQ ID NO:268 or fragments thereof may be used to diagnose Ewing sarcoma, rhabdomyosarcoma, breast cancer or germ cell tumors. In such techniques, the level of the protein of SEQ ID NO:268 in an ill individual is measured using techniques such as those described herein. The level of the protein of SEQ ID NO:268 in the ill individual is compared to the level in normal individuals. An elevated level or decreased level of the protein of SEQ ID NO:268 relative to normal individuals suggests that the ill individual is suffering from a defect in intercellular communication or secretion.

Another embodiment of the invention relates to compositions and methods using the protein of SEQ ID NO:268 or a fragment thereof as possible targets for vaccine-based therapies of cancer,

including Ewing sarcoma, rhabdomyosarcoma, breast cancer or germ cell tumors. In such embodiments, an antibody against against the protein of SEQ ID NO:268 or a fragment thereof is administered to an individual suffering from cancer in an amount sufficient to ameliorate or eliminate the cancer.

Protein of SEQ ID NO:399 (internal designation (160-40-1-0-H4-CS)

The protein of SEQ ID NO:399 is encoded by the cDNA of SEQ ID NO:158. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:399 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 160-40-1-0-H4-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:158 described throughout the present application also pertain to the human cDNA of clone 160-40-1-0-H4-CS. The protein of the invention is found to be expressed in testis and lungs. It is over represented in fetal brain.

The protein of SEQ ID NO:399 encoded by the cDNA of SEQ ID NO:158 is homologous to proteins of the Phosphatic Acid Phosphatase type 2 (PAP2) superfamily (Stukey J. and Carman G.M., Protein Sci 1997;6 :469-472, the disclosure of which is incorporated herein by reference in its entirety). Three variants of human PAP, i.e. PAP-alpha 2 (W79285) and its alternatively spliced form PAP-alpha 1 (W79284), PAP-beta (W79286) and PAP-gamma (W79287) have been identified. The protein of SEQ ID NO:399 displays a pfam characteristic domain of the PAP2 superfamily from positions 19 to 175. Accordingly, one embodiment of the present invention is a polypeptide comprising amino acid residues 19 to 175 of SEQ ID NO:399. Four membrane spanning domains are predicted from amino acid ositions 17 to, 47 to 67, 108 to 128, and 141 to 161. Accordingly, another embodiment of the present invention is a polypeptide comprising one or more of the foregoing membrane spanning domains.

Phosphatidic acid phosphatase (PAP) (also referred to as phosphatidate phosphohydrolase) is known to be an important enzyme for glycerolipid biosynthesis. In particular, PAP catalyzes the conversion of phosphatidic acid (PA) into diacylglycerol (DAG). PA and DAG are lipids involved in signal transduction and in structural membrane-lipid biosynthesis in cells, thus they represent an important regulatory point in eukariotic phospholipid metabolism. DAG is a well-studied lipid second messenger which is essential for the activation of protein kinase C (Kent; Anal. Rev. Biochem. ; 64 : 315-343;1995; whereas PA itself is also a lipid messenger implicated in various signaling pathways such as NADPH oxidase activation and calcium mobilization (English; Cell Signal.; 8:341-347 ;1996, the disclosure of which is incorporated herein by reference in its entirety). The regulation of PAP activity can therefore affect the balance of divergent signaling processes that the cell receives in terms of PA and DAG (Brindley et al.; Chem.Phys. Lipids 80:45-57 ; 1996, the disclosure of which is incorporated herein by reference in its entirety).

PAP exists in at least two isoforms, one of which (PAP1) is presumed to be cytosolic and membrane associated and the other (PAP2) to be an integral membrane protein (Leung D.W., Tompkins C.K., White T. ; DNA Cell Biol.17 : 377-385 (1998)). The protein of the invention has 180 amino-acids and four predicted membrane-spanning segments, so is presumed to be an integral membrane protein.

The protein of SEQ ID NO:399 is encoded by a cDNA that has homology to many forms of alternative splicing of PAP2 genes. For example, the protein of SEQ ID NO:399 has 29% homology with human phosphatidic acid phosphohydrolase type-2C protein. The protein of SEQ ID NO:399 also has 40% homology with human phosphatidic acid phosphatase 2B protein. In addition, the protein of SEQ ID NO:399 has 33% homology with human type 2 phosphatidic acid phosphatase alpha-2 protein. PAP2-alpha2 is one of the two isoforms with PAP2-alpha1, presumed to be alternative splice variants from a single gene.

Northern analysis has shown that PAP2-alpha mRNA expression was suppressed in several tumor tissues, indicating that PAP-2 may act as a tumor suppressor. The relationship of PAP and tumor suppression is further evidenced in findings that PAP activity is lower in fibroblast cell lines transformed with either the *ras* or *fps* oncogene than in the parental rat1 cell line (Brindley et al ; Chem. Phys. Lipids 80 : 45-57 ;1996, the disclosure of which is incorporated herein by reference in its entirety). As discussed above, a decrease in PAP activity in transformed cells correlates with a concomitant increase in PA concentration. Moreover, elevated PAP activity and lower levels of PA have been observed in contact-inhibited fibroblasts relative to proliferating and transformed fibroblasts (Brindley et al ; Chem. Phys. Lipids 80: 45-57;1996, the disclosure of which is incorporated herein by reference in its entirety). Therefore, the protein of SEQ ID NO:399 or fragments thereof may be used to decrease cell division and as such can provide a useful tool in treating cancer. Subsequent analysis of colon tumor tissue derived from four donors confirmed lower expression of PAP2-alpha than in matching normal colon tissue. Considering these data and previous demonstrations that certain transformed cell lines have lower PAP activity, human PAP cDNAs may be used for gene therapy for certain tumors (Leung D.W., Tompkins C.K., White T. ; DNA Cell Biol.17 : 377-385 (1998), the disclosure of which is incorporated herein by reference in its entirety). Accordingly, one embodiment of the present invention is the use of the protein of SEQ ID NO:399 or a fragment thereof as a tumor suppressor. For example, a nucleic acid expressing the protein of SEQ ID NO:399 or a fragment thereof may be introduced into an individual suffering from cancer in order to ameliorate or eliminate the cancer. In fact, nucleic acids encoding human phosphatidic acid phosphatases have been used to regulate levels of lipid cellular mediators and in gene therapy of e.g. cancer (PCT publication WO98/46730, the disclosure of which is incorporated herein by reference in its entirety).

In another embodiment of the present invention, the protein of SEQ ID NO:399 or a fragment thereof can be used to control the balance of lipid mediators of cellular activation and

signal transduction. The protein of the invention has 33% homology with human phosphatidic acid phosphatase 2A protein. PAP2A is an integral membrane glycoprotein at the cell surface that plays an active role in the hydrolysis and uptake of lipids from the extracellular space (Roberts RZ, Morris AJ; Biochim Biophys Acta 2000 Aug 24;1487(1):33-49, the disclosure of which is incorporated herein by reference in its entirety). Accordingly, the level or activity of the protein of SEQ ID NO:399 may be modulated to influence the rate or extent of hydrolysis and uptake of lipids from the extracellular space using methods such as those described herein.

In another embodiment of the present invention, the protein of SEQ ID NO:399 can be used to counterbalance the inflammatory response. PA has been implicated in cytokine induced inflammatory responses (Bursten et al; Circ. Shok 44: 14-29, 1994; Abraham et al; J. Exp. Med. 181: 569-575, 1995; Rice et al; PNAS 91: 3857-3861, 1994; Leung et al; PNAS 92: 4813-4817, 1995, the disclosures of which are incorporated herein by reference in their entireties) and the modulation of numerous protein kinases involved in signal transduction (English et al ; Chem. Phys. Lipids 80: 117-132, 1996, the disclosure of which is incorporated herein by reference in its entirety). In addition, a nucleic acid encoding the protein of SEQ ID NO:399 or a fragment thereof may be used to counterbalance the inflammatory response from cytokine stimulation through degradation of excess amount of PA in cells or to treat or ameliorate inflammatory diseases.

The gene encoding the protein of SEQ ID NO:399 or a fragment thereof can also be used in gene therapy for the treatment of obesity associated with diabetes. PAP activity is decreased in the livers and hearts of the grossly obese and insulin resistant JCR:LA corpulent rat compared to the control lean phenotype (Brindley et al ; Chem. Phys. Lipids 80 : 45-57 ;1996, the disclosure of which is incorporated herein by reference in its entirety). The protein of the invention therefore can provide an important tool for the treatment of obesity associated with diabetes.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:399 , fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition, such as those listed above, in an individual. In such embodiments, the protein of SEQ ID NO:399 , or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ ID NO:399, including glycerolipid biosynthesis, conversion of phasphatidic acid into diacylglycerol, signal transduction, membrane-lipid biosynthesis, activation of protein kinase C, NADPH oxidase activation, calcium mobilization, cell division, production of diacylglycerol, monoacylglycerol, ceramide or sphingosine, modulation of the inflammatory response or dephosphorylation of a substrate such as lysophasphatidic acid, ceramide 1-phosphate, or sphingosine 1-phosphate, or treatment or amelioration of obesity associated with diabetes. The protein of SEQ ID NO:399 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:399 or a fragment thereof may be administered to the individual.

Alternatively, an agent which increases the activity of the protein of SEQ ID NO:399 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:399 or a cell or preparation containing the protein of SEQ ID NO:399 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:399 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:399 may be identified by contacting the protein of SEQ ID NO:399 or a cell or preparation containing the protein of SEQ ID NO:399 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably brain, or to distinguish between two or more possible sources of a tissue sample on the basis of the level of the protein of SEQ ID NO:399 in the sample. For example, the protein of SEQ ID NO:399 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples,differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a tissue sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from brain or tissues other than brain to determine whether the test sample is from brain. Alternatively, the level of the protein of SEQ ID NO:399 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:399 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other technques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from brain or tissues other than brain to determine whether the test sample is from brain.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:399 , including using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:399 or a fragment thereof may be fixed to a solid support, such as a chromatograpy matrix. A prepartation containing cells expressing the protein of SEQ ID NO:399 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The

support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:399 or a fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:399. In some embodiments, the protein of SEQ ID NO:399 or fragments thereof may be used to diagnose cancer. In such techniques, the level of the protein of SEQ ID NO:399 in an ill individual is measured using techniques such as those described herein. The level of the protein of SEQ ID NO:399 in the ill individual is compared to the level in normal individuals. An elevated level or decreased level of the protein of SEQ ID NO:399 relative to normal individuals suggests that the ill individual may suffer from cancer or be predisposed to getting cancer in the future.

In another embodiment, the present invention relates to methods of preparing a PAP protein of SEQ ID NO:399 comprising the steps of (i) transforming a host cell with an expression vector comprising a polynucleotide encoding SEQ ID NO:399, (ii) culturing the transformed host cells which express the protein and (iii) isolating the protein. The present invention also relates to a method of dephosphorylating a substrate comprising contacting the substrate with an effective amount of isolated protein of SEQ ID NO:399 or a fragment thereof such that the protein catalyzes the dephosphorylation of the substrate. It is further provided that this method occurs *in vitro*, and comprises a step of isolating the dephosphorylated substrate. Additionally, the method can occur *in vivo*, and is effected by the administration of the protein of the invention (or part of it) to a mammal in need thereof.

Protein of SEQ ID NOs:258 and 262 (internal designations 110-007-1-0-C7-CS, 116-055-1-0-A3-CS):

The protein of SEQ ID NO:258 is encoded by the cDNA of SEQ ID NO:17. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:258 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 110-007-1-0-C7-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:17 described throughout the present application also pertain to the human cDNA of clone 110-007-1-0-C7-CS. The protein of SEQ ID NO:258 shows homologies to two high affinity IgE receptor-like proteins (IGER) with GENESEQP accession numbers W96745 and W41056, the disclosures of which are incorporated herein by reference in their entireties. The protein of SEQ ID NO:258 is expressed in liver and testis. The protein of SEQ ID NO:262, encoded by SEQ ID NO:21, is a variant of the protein of SEQ ID NO:258 and shares all the potential uses and functions described herein. This protein and cDNA share all of the characteristics and uses of the clone, and product thereof, 116-055-1-0-A3-CS).

Like the two high affinity IgE receptor-like proteins, the protein of the invention contains four transmembrane spanning domains of 20 amino acids, between amino acids 53-73, 79-99, 121-141 and 158-178, respectively. The protein of SEQ ID NO:258 crosses the plasma membrane four times forming two small extracellular loops and has both the N- or C- terminals in the cytoplasm. Moreover, the protein of the invention contains a signal peptide (cleavage site at position 21).

The predicted structure of the protein of SEQ ID NO:258 demonstrates the relationship of this protein to FcεRIβ and CDC20 antigen and provides evidence for a family of 4-transmembrane spanning proteins. The conservation of amino acids between all three proteins is highest in the four transmembrane domains. While greater divergence exists in the hydrophilic amino and carboxyl termini, several amino acids within these regions are conserved such as the presence of 4 prolines in the amino terminus of all three proteins. In addition, two cysteine residues (position 147 and 156) are present in the second extracellular domain between TM3 and TM4. This suggests that inter- or intra-molecular di-sulfite bonds in this domain are present in all three proteins.

FcεRI, is part of a tetrameric receptor complex consisting of an α chain, a β chain and two γ chains (Kinet et al. Proc Natl. Acad. Sci. USA, 15: 6483-6487 (1988), the disclosure of which is incorporated herein by reference in its entirety). Together, they mediate interaction with IgE-bound antigens leading to dramatic cellular responses, such as the massive degranulations of mast cells. The β subunit is a 4-transmembrane protein with both the amino and carboxyl termini residing in the cytoplasm.

Chromosome mapping localized cDNA of SEQ ID NO:17 to chromosome 11q12, the location of the CD20 gene. However, the murine FcεRIβ and Ly-44 (the murine equivalent of CD20) are both located in the same position in mouse in chromosome 19 (Teder, T.F. et al., J. Immunol. 141:4388-4394 (1988), Clark E.A. and Lane, J.L. Annu. Rev. Immunol. 9:97-127 (1991), the disclosures of which are incorporated herein by reference in their entireties). Therefore, the three genes are believed to have been originated and evolved from the same locus, further supporting the proposition that they are members of the same family of related proteins.

On the basis of the foregoing information, it is believed that the protein of SEQ ID NO:258 is a high affinity immunoglobulin E receptor-like protein.

Atopic diseases, which include allergy, asthma, atopic dermatitis (or eczema) and allergic rhinitis are generally defined as a disorder of Immunoglobulin E (IgE) responses to common antigens, such as pollen or house dust mites. It is frequently detected by either elevated total serum IgE levels, antigen specific IgE response or positive skin tests to common allergens. In principle, atopy can result from dysregulation of any part of the pathway which begins with antigen exposure and IgE response to the interaction of IgE with its receptor on mast cell, the high affinity Fc receptor FcεRI, and the subsenquent cellular activation mediated by that ligand-receptor engagement (Ravetch, Nature Genetics, 7: 117-118 (1994), the disclosure of which is incorporated herein by reference in its entirety).

Accordingly, the protein of SEQ ID NO:258 or fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity may administered to an individual in whom it is desired to increase or decrease the activity of the protein of SEQ ID NO:258. In particular, the protein of SEQ ID NO:258 or fragment thereof may be administered to an individual in whom it is desired to regulate the extent of the IgE response. In such methods, the protein of SEQ ID NO:258 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:258 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:258 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:258 or a cell or preparation containing the protein of SEQ ID NO:258 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

The protein of SEQ ID NO:258 or fragments thereof may also be used to identify genes or polypeptides that may play a rôle in IgE responses or atopic disease. In particular, binding partners for the protein of SEQ ID NO:258 or the genes encoding such binding partners may be identified using a variety of techniques familiar to those skilled in the art, including the techniques described herein.

The protein of SEQ ID NO:258 or the polynucleotide encoding the protein of SEQ ID NO:258 may also be used to diagnose hereditary atopy. In particular, the level of the protein of SEQ ID NO:258 may be determined in a test individual using methods such as those described herein and compared to the levels of normal individuals and individuals suffering from hereditary atopy to determine whether the test individual is suffering from or at risk of suffering hereditary atopy. Alternatively, a nucleic acid sample may be obtained from a test individual and analyzed to determine whether it contains a level of RNA encoding the protein of SEQ ID NO:258 which is associated with hereditary atopy or a mutation in the gene encoding the protein of SEQ ID NO:258 which is associated with hereditary atopy. For example, a nucleic acid sample from the test individual may be contacted with a nucleic acid probe comprising the nucleic acid encoding the protein of SEQ ID NO:258 or a fragment thereof to determine the RNA level or whether the individual has a mutation associated with hereditary atopy. The probe may be either DNA, including cDNA or genomic DNA, or the probe may be RNA. Any of the methods familiar to those skilled in the art may be used in these diagnostic methods, including the methods described herein. For example, the presence of a mutation associated with hereditary atopy can be determined using methods generally known in the art, such as but not limited to PCR, sequencing or mini sequencing as described in the method of Yamamoto et al. (Biochem. Biophys. Res. Comm., 182:507 (1992), the disclosure of which is incorporated by reference herein in its entirety).

The protein of SEQ ID NO:258 can also be used to characterize the induction of expression of FcεRI and the particular function of FcεRIβ. As such, the protein of the invention can be useful in, for example, the design of drugs that block or inhibit induction or activity of FcεRI, thereby treating atopic diseases. In particular, test agents which block or inhibit induction or activity may be identified using the methods described herein.

In an other embodiment, the protein of SEQ ID NO:258 can be employed in the preparation of antibodies, such as monoclonal antibodies, according to methods known in the art, including those described herein. The antibodies can be used to block or mimic ligand binding to the receptor comprising the protein of the invention or other receptors, such as but not limited to FcεRI. The antibodies can also be used to isolate the protein of SEQ ID NO:258 or cells which express the protein of SEQ ID NO:258 using methods such as those described herein. For example, the antibodies may be used to measure the presence of cells containing the protein of SEQ ID NO:258 (including but not limited to hematopoietic cells) in a sample. For example, the method comprises contacting the sample with the antibody under conditions sufficient for the antibody to bind to the protein of SEQ ID NO:258 and detecting the presence of bound antibody using methods known in the art, including those described herein.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably liver and testis, or to distinguish between two or more possible sources of a tissue sample on the basis of the level of the protein of SEQ ID NO:258 in the sample. For example, the protein of SEQ ID NO:258 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples,differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a tissue sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from liver or testis or tissues other than liver or testis to determine whether the test sample is from liver or testis. Alternatively, the level of the protein of SEQ ID NO:258 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:258 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other technques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from liver or testis or tissues other than liver or testis to determine whether the test sample is from liver or testis.

Protein of SEQ ID NO:279 (internal designation 160-58-3-0-H3-CS)

The protein of SEQ ID NO:279 is encoded by the cDNA of SEQ ID NO:38. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:279 described throughout the present application also pertain to the polypeptide encoded by a nucleic acid included in clone 160-58-3-0-H3-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:38 described throughout the present application also pertain to the nucleic acid included in clone 160-58-3-0-H3-CS.

The protein of SEQ ID NO:279 is encoded by a nucleic acid of 1330 nucleotides with an ORF between nt 198 to 998 yielding a 267 amino acid protein. The protein is a polymorphic variant of the sequence (SP:P01210) for proenkephalin A precursor (contains Met- and Leu- enkephalins). It has a signal peptide spanning 24 amino acid and 2 signature motifs for vertebrate endogenous opioid neuropeptides and endogenous opioid neuropeptide precursors. PSORT gives a predicted extracellular localization, including the cell wall (66.7%). The protein of SEQ ID NO:279 is primarily distributed the fetal brain, although expression in other tissues has also been shown (see below). The polymorphic variation is found at amino acid position 75 (E→D, a conservative amino acid change). After signal peptide cleavage (amino acid 47 to 267; 220 amino acid), the protein still contains the polymorphic variation, which is now at amino acid position 29. This does not change any of the sequence of the different enkephalins that result after cleavage of this precursor protein. In addition, the polymorphism is 25 amino acids away from the first cleavage site on the amino terminal side. This is unlikely to change the secondary structure of the actual cleavage site.

PCT publication WO9606863-A1, the disclosure of which is incorporated herein by reference in its entirety, discloses a protein having high homology with the protein of SEQ ID NO:279. Accordingly, the protein of SEQ ID NO:279 is believed to be an enkephalin. Met- and Leu- enkephalins compete with and mimic the effects of opiate drugs. These two pentapeptides with potent opiate agonist activity in bioassay systems were originally identified by Hughes et al (Nature, 258, 577-580, 1975). The natural ligands for opiate receptors, which differ only in their COOH terminal amino acid, were named Met- and Leu-enkephalin to reflect their origin from the brain. Peptides containing these sequences are termed opiate or opioid peptides. Enkephalins are widely distributed throughout the central nervous system in enkephalinergic neuronal networks, and also exist in the peripheral nervous system, for example in autonomic ganglia. Data, largely circumstantial, suggest wide-ranging involvement of endogenous opioids for example in the modulation of pain perception, in mood and behaviour, learning and memory, responses to stress, diverse neuroendocrine functions, immune regulation and cardiovascular and respiratory function.

Met-enkephalin enhances the immune reaction in patients with cancer or AIDS. It can bind opoid receptors present in peripheral inflamed tissues to mediate an analgesic effect.

After exogenous administration of the different enkephalins, several immunologic functions are affected, including antibody production, NK cell activity against tumors and viral infections,

macrophage and polymorphonuclear leukocyte functions, graft rejections, and mitogen-stimulated lymphocyte proliferation. The effects can be bi-directional, where low concentrations enhance, and high concentrations inhibit the same immune function. Thus, enkephalins are modulators of immune reactions.

These opioid neuropeptides are released by post-translational proteolytic processing of precursor proteins. These multivalent precursor proteins (polyprotein) consist of a signal sequence followed by a conserved region of about 50 residues, a variable length region and the sequence of the various neuropeptides. The preproenkephalin A (gene PENK) is processed to produce the following peptides which include Met-enkephalin (6 copies, 2 of which are extended) and Leu-enkephalin:

Signal peptide 1-24

Peptide 100-104 Met-enkephalin 1

Peptide 107-111 Met-enkephalin 2

Peptide 136-140 Met-enkephalin 3

Peptide 186-193 Met-enkephalin-arg-gly-leu

Peptide 210-214 Met-enkephalin 4

Peptide 230-234 Leu-enkephalin

Peptide 261-267 Met-enkephalin-arg-phe

The conserved region in the N-termini of these precursors contains six cysteines that are probably involved in disulfide bonds. This region could also be important for the processing of the neuropeptides.

The precursor protein does have the potential to be differentially cleaved into multiple extended enkephalin and non-enkephalin-containing peptides, the functions of which are largely unknown; however, in some cases it has been shown that extended enkephalin-containing peptides have enhanced opiate activity. Another peptide, enkelytin, is produced that exhibits anti-bacterial activity (see below).

There is a growing body of evidence that proenkephalin exists largely independently of free enkephalin peptides in a number of tissues and cell types including astrocytes (Melner et al, EMBO J, 9, 791-796, 1990; Spruce et al, EMBO J 9, 1787-1795, 1990, the disclosures of which are incoporated herein by reference in their entireties), and is released from these cells in an unprocessed form (Batter et al, Brain Res. 563, 28-32, 1991, the disclosure of which is incorporated herein by reference in its entirety). There is evidence in some cases that processing enzymes are co-released along with the unprocessed precursor which suggests that extracellular cleavage may occur (Vilijn et al, J. Neurochem. 53, 1487-1493, 1989). Even if biological activity is signalled through binding of the small peptide products to cell surface receptors, the regulation of this activity may be mediated through the precursor, and it is also possible that the unprocessed precursor has an additional intracellular role of its own.

This protein was originally described to be present in various brain regions, most notably in the striatum as well as in neuroendocrine tissues, the pituitary and adrenal gland. It is also expressed in a variety of immune cells, including ConA-stimulated CD4 Tlymphocytes, CD4 thymocytes, B lymphocytes, as well as T cell lines, macrophages and mast cells. Expression has been reported in the reproductive system, heart and many developing tissues during gestation and early postnatal period Because of this, it has been postulated that these peptides play a role in cell or tissue growth and differentiation. For example, endogenous enkephalins induced in thymocytes modulate their own expression and function to inhibit the proliferation of activated thymocytes.

Enkephalin peptides are abundant in adrenal medulla and can be released by neurotransmitters specific for that tissue. Enkephalins have also been found to be abundant in human phaeochromocytoma, a tumour derived from the adrenal medulla. The RNA from this tumour contains a high level of enkephalin mRNA sequences as demonstrated by cell-free translation studies.

Enkephalins function as opiate receptors are classified as delta, kappa and mu. A study by Lord et al (Nature, 267, 495-499, 1977) compared the activity of morphine and enkephalins in bioassay systems, and found that enkephalins bound predominantly to delta receptors. Subsequent studies have revealed homology of these receptors to other receptor families, including the immunoglobulin superfamily member OBCAM (Schofield et al, EMBO J 8, 489-495, 1989, the disclosure of which is incorporated herein by reference in its entirety) and somatostatin receptors (PCT publication WO96/06863, the disclosure of which is incorporated herein by reference in its entirety). This would explain the reported opioid binding properties of the former. Because of the latter's homology to opiate receptors, it would also be expected to bind opioid receptor ligands. The recognition of opioid peptides by other non-opiate related receptors implies that these peptides may exert other as yet unknown functions.

Enkephalins are also involved in apoptosis. Apoptosis is the morphologically distinct process of controlled cell death which balances the process of cell production by mitosis. A molecular connection between control of cell production and cell elimination has now been established, including the roles of c-myc and p53 in the pathways mediating apoptotic cell death. It has been proposed that all mammalian cells may be programmed to die by default in the absence of continuous signalling from neighboring cells. However, the acquisition of a survival advantage which prevents a single cell from activating its suicide program in response to levels of genetic damage associated with common environmental insults could theoretically be an initiating event in oncogenesis since it would favor the persistence of potentially tumorigenic mutations. Alternatively, inappropriate activation of survival pathways might lead to overriding the intrinsic death program and promote tumorigenesis at early and late stages. A particularly potent oncogenic pathway would be one which both promoted and tolerated genetic damage and helped a cell overcome its need for extracellular survival signals. Approximately 50% of human tumors possess

normal p53 function. Thus, additional pathways or molecules which inappropriately repress apoptosis in human tumours remain to be identified. Opioid-like molecules could be involved in such a pathway.

There are published reports that pathways which include opioid-like molecules participate in regulating the equilibrium between cell death and survival. For example, morphine inhibits cell survival in the developing cerebellum (Hauser et al, Exp. Neurol, 130, 95-105, 1994, the disclosure of which is incorporated herein by reference in its entirety) and induces apoptosis in thymocytes (Fuchs and Pruett, J. Pharmacol. Exp. Ther. 266, 417-423, 1993, the disclosure of which is incorporated herein by reference in its entirety).

In a series of experiments (PCT publication WO 96/06863), it has been found that proenkephalin and/or its proteolytic products act as extracellular and/or cell surface membrane bound factors which modulate cell survival in transformed cells a) upon deprivation of exogenous survival factors, and b) following genotoxic injury and/or stress when exogenous survival factors are non-limiting. The receptor(s) to which these factor(s) bind, which are most likely to exist on the cell surface are related, or possibly identical, to one or more members of the opioid receptor family.

Opioid-like receptor types or subtypes can mediate survival or death; receptor (s) whichmediate death appear to be coupled to those which mediate survival. Natural ligands for these receptors are likely to be products of the opioid precursor genes, although natural ligands could include cytokines which mimic their effect. Tumour cells are more sensitive to antagonism of opioid-like receptor-mediated survival, and to stimulation of opioid-like receptor-mediated death, than non-transformed cells. The induction of cell cycle arrest enhances the sensitivity of tumour cells to thesemanipulations. (Enhanced sensitivity of tumour cells to these manipulations is induced by their synchronisation within the cell cycle.

Cytoplasmic proenkephalin and/or its proteolytic products act as general repressors of apoptosis. Agents which, if coupled to appropriate internalisation agents, would antagonise cytoplasmic proenkephalin would therefore be of use in the induction of apoptosis in non-transformed as well as transformed cells, particularly in combination with sublethal doses of known apoptosis-inducing agents.

The repression of apoptosis mediated through cytoplasmic proenkephalin is activated at high cell density predominantly by nondiffusable factors. Inhibition of proenkephalin or its products as described above would therefore be potentiated if agents were used in combination for example with neutralising antibodies to integrins (such as the antibody 23C6- Bates et al., J. Cell Biol. 125 403-415, 1994) to reduce exogenous survival signaling and simulate low density.

Proenkephalin targeted to the cell nucleus induces apoptotic death, which is inhibited by the overexpression of large T antigen and is at least partly mediated through p53. Tumors which retain wild-type p53 function are therefore a particular target for apoptosis induction by agents which

increase the levels of proenkephalin, or its derivatives, within the nucleus or which mimic the function of nuclear proenkephalin or its derivatives.

Accordingly, the protein of SEQ ID NO:279, fragments thereof, or nucleic acids encoding the protein of SEQ ID NO:279 may be used to modulate a biochemical pathway in which products of opioid peptide precursor genes participate. In some embodiments, antibodies or other agents which reduce the level or activity of the protein of SEQ ID NO:279 or fragments thereof may be used to induce apoptosis in cells. The agents preferably neutralize the protein of SEQ ID NO:279 or its proteolytic derivatives, increase the level of, activate or mimic nuclear proenkephalin, or act as an antagonist to receptors related or identical to the delta and kappa opioid receptors. In some embodiments, the agent may be a neutralizing monoclonal antibody against the protein of SEQ ID NO:279 or a fragment thereof. The agent may also be a fragment or allelic form of one of these antibodies. A cytoplasmic anchor, or a nuclear localization signal may also be included in the agent. In some embodiments, the agent is able to modulate a biochemical pathway in a cell in which products of opioid peptide precursor genes participate in order to induce apoptosis. The agents can be used for the treatment of cancer or for inducing apoptosis in lens cells following a cataract operation. In some embodiments, the agents promote apoptosis of proliferating cells with less, or no, effect on normal mature cell types. The agents may be administered in combination with a genotoxic or cell cycle arrest agent. Alternatively, the agent may be complexed with a chemotherapeutic, irradiation or cell cycle arrest (synchronization agent).

Accordingly, the invention provides a means of inducing apoptosis in cells which comprises modifying a biological pathway of a cell in which a product of an opioid precursor gene participates in such a way that apoptosis is induced. Modification of the pathway is suitably effected by adminstration of an appropriate agent. In particular, the present invention provides an agent for use in inducing apoptosis in cells, said agent comprising an agent able to neutralise proenkephalin or its proteolytic derivatives; an agent which increases the level of nuclear proenkephalin and/or its derivatives, or which activates or mimics them an agent which acts as an antagonist at receptor(s) related or identical to the delta opioid receptor, or an agent which acts as an agonist at receptor(s) related or identical to the kappa opioid receptor.

A subset of such agents are agents able to neutralise proenkephalin or its proteolytic derivatives, or an agent which acts as an antagonist at receptor(s) related or identical to the delta opioid receptor, or an agent which acts as an agonist at receptor(s) related or identical to the kappa opioid receptor.

In some embodiments, the agent may be administered to the cell surface whereupon the survival effects of extracellular and/or cell surface membrane bound proenkephalin or its proteolytic derivatives is neutralised causing the cell to become apoptotic. Alternatively, an agent able to neutralise proenkephalin or its proteolytic derivatives may be coupled to an internalisation peptide and a cytoplasmic anchor. Such an assembly will remain in the cytoplasm of the cell, antogonising

cytoplasmic proenkephalin and/or its proteolytic products and thus neutralising the apoptosis repressor effect of these molecules.

Enkephalins also have anti-bacterial activity. During processing of the proenkephalin-A, the maturation in the adrenal medullary chromaffin cell starts with the removal of the carboxy-terminal end (proenkephalin-A-derived peptide or $PEAP_{209-239}$) (Y. Goumon, K. Lugardon, B. Kieffer et al. J. Biol. Chem. 273:29847-29856, 1998, the disclosure of which is incorporated herein by reference in its entirety). The peptide enkelytin was identified as corresponding to bisphosphorylated $PEAP_{209-237}$, and possesses antibacterial activity including *Staphylococcus aureus* and other gram-positive bacteria such as *Micrococcus luteus* and *Bacillus megaterium* (0.2-0.4 μM range). There is no ability to affect gram-negative bacteria (E. coli strain D22, D31, 663 and T13773) growth, nor is there any hemolytic activity. The activity of this peptide is specific – shorter versions of the peptide (209-220, 224-237, 230-237, 233-237) or non-phosphorylated $PEAP_{209-239}$ exhibited little to no bacterial growth inhibiting activity.

Bovine periarthritis abscess fluid contains different forms of PEAP (72-237/239; 80-237/239) as identified by immunoreactivity and confirmed by sequence analysis. These peptides have activity against *M. luteus*, but are less active than enkelytin (5 versus 0.2 μM). These PEAP constitute a pool of precursors which have to be processed, during infection, to provide active enkelytin. Presence of a PEAP at a molecular mass corresponding to that of $PEAP_{209-237}$ was detected as well. PEAPs ($PEAP_{202-238}$ and $PEAP_{206-237}$) have also been detected in wound fluids, including bovine post-caesarean abscess in the subcutaneous lining, and an abscess induced by subcutaneous injection of complete Freund's adjuvant. Therefore, these peptides are present in wound fluids along with other known antibacterial peptides (defensins, bactenecins). The concentrations were in a range similar to that found to be active *in vitro* (0.5-1 μM). The PEAPs have also been detected in secretions from human polymorphonuclear neutrophils.

The PEAP209-230 and enkelytin are secreted from cultured chromaffin cells following stimulation. This suggests that these two peptides are co-released with catecholamines in stress situations and may therefore play an important role in defense mechanisms.

Co-release of met-enkephalin and enkelytin would represent a unified neuroimmune protective response to stress situations that may be accompanied with infectious diseases. This would provide a highly beneficial survival strategy at the very begninning of proinflammatory processes. This protein would therefore play an important role in host defense against microbial infections, especially those involving gram positive bacteria. Due to their nonspecific activity on membranes, the antibacterial peptides possess cytotoxic activities and may not only play a role in antimicrobial defense, but also in inflammatory processes, possibly in wound repair.

The protein of SEQ ID NO:279, peptides derived by cleavage thereof or fragments thereof could be used as antibacterial agents in creams/ointments/solutions, presoaked bandages, or dermal-type patches for external applications. Alternatively, the protein of SEQ ID NO:279, peptides

derived by cleavage thereof, or fragments thereof may be used in injections (intravenously, subcutaneously or intra-peritoneally). This is useful for wound repair, burn healing, post-operative recovery management.

Alternatively, the protein of SEQ ID NO:279, peptides derived by cleavage thereof, or fragments thereof, may be incorporated into disinfectant solutions used for cleaning surfaces such as in the the house (kitchen, bathroom) or in the office (desktops, phones, computer keyboards and mouse). Other applications are as additives in mouthwash or handi-popup wipes.

Altered levels of enkephalins may produce psychological disease. Konig et al (Nature, 383, 535-538, 1996, the disclosure of which is incorporated herein by reference in its entirety) used a genetic approach to study the role of the mammalian opioid system. They disrupted the pre-proenkephalin gene using homologous recombination in embryonic stem cells to generate enkephalin-deficient mice. Mutant enk -/- animals are healthy, fertile, and care for their offspring, but display significant behavioral abnormalities. Mice with the enk -/- genotype are more anxious and males display increased offensive aggressiveness. Mutant animals show marked differences from controls in supraspinal, but not in spinal, responses to painful stimuli. These enk -/- mice do however exhibit normal stress-induced analgesia. Therefore, enkephalins modulate responses to painful stimuli. Thus, genetic factors may contribute significantly to the experience of pain. This study clearly indicates the importance of enkephalins in pain perception, anxiety and aggressiveness.

Interestingly, the PENK gene is localized on 8q23-q24, the same locus on which are found genes related to epilepsy and spastic paraplegia, disorders related to brain dysfunction.

Accordingly, the protein of SEQ ID NO:279 or fragments thereof may be used for the treatment of psychological disorders, especially those involving distortion in the perception of pain, aggressiveness, or anxiety. This would include drug addiction, different types of phobias, panic attacks, schizophrenia, bi-polar, anorexia nervosa, chronic pain disorders, post-traumatic events. post-operative pain management.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:279, fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition in an individual. For example, the condition may be cancer, a condition resulting from increased or decreased cellular proliferation, bacterial infection, conditions resulting from abnormal immune responses, psychological disease or any of the conditions listed above. In such embodiments, the protein of SEQ ID NO:279, or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ ID NO:279. The protein of SEQ ID NO:279 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:279 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the

activity of the protein of SEQ ID NO:279 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:279 or a cell or preparation containing the protein of SEQ ID NO:279 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:279 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:279 may be identified by contacting the protein of SEQ ID NO:279 or a cell or preparation containing the protein of SEQ ID NO:279 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify the source of a sample as, for example, fetal brain, or to distinguish between two or more possible sources of a sample on the basis of the level of the protein of SEQ ID NO:279 in the sample. For example, the protein of SEQ ID NO:279 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from fetal brain or tissues other than fetal brain to determine whether the test sample is from fetal brain. Alternatively, the level of the protein of SEQ ID NO:279 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:279 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other technques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from fetal brain or tissues other than fetal brain to determine whether the test sample is from fetal brain.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:279, including using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:279 or a fragment thereof may be fixed to a solid support, such as a chromatograpy matrix. A preparation containing cells expressing the protein of SEQ ID NO:279 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The

support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:279 or a fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:279. In such techniques, the level of the protein of SEQ ID NO:279 in an ill individual is measured using techniques such as those described herein. The level of the protein of SEQ ID NO:279 in the ill individual is compared to the level in normal individuals to determine whether the individual has a level of the protein of SEQ ID NO:279 which is associated with disease.

Protein of SEQ ID NO: 293 (internal designation 181-16-1-0-G7-CS)

The protein of SEQ ID NO: 293 has a high degree of homology with HSPC163 (Genbank accession number AF161512), the protein encoded by gene no: 93 (PCT/US99/17130) and the human cornichon protein TGAM77. SEQ ID NO: 293 is overexpressed in cancerous prostate, fetal brain and fetal kidney.

The gene HSPC163 is one of three hundred cDNAs obtained from CD34+ hematopoietic stem / progenitor cell (HSPC) library (obtained from umbilical cord blood and adult bone marrow). HSPC163 has also been in identified in five hematopoietic cell lines: NB4 (granulocytic), HL60 (granulocytic), U937 (monocytic), K562 (erythro-megakaryocytic), and Jurkat (T lymphocytic). These cell lines represent the distinct lineages of hematopoietic cells.

The polypeptide of gene no: 93 has been determined to have two transmembrane domains and a short cytoplasmic tail. Based upon these characteristics, it is believed that the protein product of gene no: 93 shares structural similarity to type IIIa membrane proteins. This gene is expressed primarily in activated T-cells and to a lesser extent in endometrial tumor, T cell helper II cells, microvascular endothelial cells, Raji cells treated with cyclohexamide and umbilical vein endothelial cells. The expression pattern of gene no: 93, indicates a role in regulating the proliferation, survival, differentiation, and/or activation of hematopoietic cell lineages, including blood stem cells. The gene product appears to be involved in the regulation of cytokine production, antigen presentation, and other immune processes, suggesting a usefulness in boosting the immune system. The translation product of this gene has high homology to the human TGAM77 and mouse cornichon proteins.

TGAM77 was identified as a gene involved in early phase of T-cell activation in response to alloantigens. Twenty four hours after T-cell allostimulation, RNA expression of TGAM77 is significantly increased. TGAM77 has been designated as a T-cell growth associated molecule. TGAM77 is a human homolog of cornichon (cni) protein of the fruit fly Drosophila.

Cornichon was demonstrated to be involved in carefully orchestrated signaling events during Drosophila oogenesis establishing an asymmetric pattern in the oocyte as a prerequisite for

correct embryogenesis. Cornichon signaling functions in concert with two other proteins. The function of all three genes in an EGF-like signaling pathway appears to direct the formation of a correctly polarized microtubule cytoskeleton, which is thought to be the basis for the correct spatial localization of other singaling molecules essential for oocyte polarization, asymmetric movement of the nucleus, and embryo differentiation.

The subject invention provides the amino acid sequence of SEQ ID NO: 293 and polynucleotide sequences encoding the amino acid sequence of SEQ ID NO: 293 . In one embodiment, the polypeptides of SEQ ID NO: 293 are interchanged with the corresponding polypeptides encoded by the human cDNA of clone 181-16-1-0-G7-CS. Also included in the invention are biologically active fragments of SEQ ID NO: 293 and polynucleotide sequences encoding these biologically active fragments. "Biologically active fragments" are defined as those peptide or polypeptide fragments of SEQ ID NO: 293 which have at least one of the biological functions of the full length protein (e.g., the ability to stimulate T-cell proliferation).

The invention also provides variants of SEQ ID NO: 293 . These variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence of SEQ ID NO: 293. Variants according to the subject invention also have at least one functional or structural characteristic of SEQ ID NO: 293, such as the biological functions described above. The invention also provides biologically active fragments of the variant proteins. Unless otherwise indicated, the methods disclosed herein can be practiced utilizing SEQ ID NO: 293 or variants thereof. Likewise, the methods of the subject invention can be practiced using biologically fragments of SEQ ID NO: 293, or variants of said biologically active fragments.

Because of the redundancy of the genetic code, a variety of different DNA sequences can encode SEQ ID NO: 293 . It is well within the skill of a person trained in the art to create these alternative DNA sequences which encode proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are, thus, within the scope of the subject invention. As used herein, reference to "essentially the same" sequence refers to sequences that have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity. Fragments retaining one or more characteristic biological activity of SEQ ID NO: 293 are also included in this definition.

"Recombinant nucleotide variants" are alternate polynucleotides which encode a particular protein. They can be synthesized, for example, by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce specific restriction sites or codon usage-specific mutations, can be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic host system, respectively.

SEQ ID NO: 293 protein, and variants thereof, can be used to produce antibodies according to methods well known in the art. The antibodies can be monoclonal or polyclonal. Antibodies can

also be synthesized against fragments of SEQ ID NO: 293 as well as variants of SEQ ID NO: 293 according to known methods. The subject invention also provides antibodies which specifically bind to biologically active fragments of SEQ ID NO: 293 or biologically active fragments of SEQ ID NO: 293 variants.

The subject invention also provides for immunoassays which are used to screen for, monitor, or diagnose prostate cancer. Methods of screening for, diagnosing, identifying, or monitoring the course of prostate cancer are well known to those skilled in the art. In this aspect of the invention, immunoassays are provided which contact a biological sample (e.g., blood, serum, tissue, or biopsied tissue sample) with antibodies which specifically bind to SEQ ID NO: 293 , immunogenic fragments of SEQ ID NO: 293 , or biologically active fragments of SEQ ID NO: 293 . Immunocomplexes formed in the contacting step are then detected using an appropriately labeled detection reagent. The levels of SEQ ID NO: 293 expressed in the tested biological samples are compared to control/normal levels typically observed in the population.

Alternatively, methods which screen for, monitor, or diagnose prostate cancer may be practiced with SEQ ID NO: 293 , or fragments of SEQ ID NO: 293 , as well as nucleic acids encoding SEQ ID NO: 293 , or fragments of SEQ ID NO: 293 . In one embodiment, the polypeptide may be used as a standard/control immunoassays described above. In another embodiment, the nucleic acids encoding SEQ ID NO: 293 , or fragments of SEQ ID NO: 293 are used in hybridization assays, well known to the skilled artisan, to identify biological samples (e.g., blood, serum, tissue, or biopsied tissue sample) which contain SEQ ID NO: 293 . The levels of SEQ ID NO: 293 expressed in the tested biological samples are compared to control/normal levels typically observed in the population.

In another embodiment, SEQ ID NO: 293 , and polynucleotide sequences encoding the amino acid sequence of SEQ ID NO: 293 can be used to identify or diagnose immune disorders involving activated T-cells using standard hybridization assays.

Another aspect of the invention provides methods of immunostimulating a mammal. In this aspect of the invention, SEQ ID NO: 293 , and/or polynucleotide sequences encoding the amino acid sequence of SEQ ID NO: 293 , are introduced into T-cells according to well known methods. T-cells are, then activated by stimulation with antigen to induce the immune system of the mammal.

In another embodiment, autologous T-cells are obtained from an individual. SEQ ID NO: 293 , biologically active fragments thereof, and/or polynucleotide sequences encoding the amino acid sequence, or biologically active fragments, of SEQ ID NO: 293 , are introduced into these autologous T-cells according to well known methods. The T-cells are expanded and reintroduced into the individual from which the T-cells were obtained. See, for example U.S. Patent Nos. 5,192,537 and 5,766,920 , hereby incorporated by reference in their entirety.

In another embodiment of the subject invention, polynucleotides and polypeptides encoding SEQ ID NO: 293 , can be used to expand stem cells, committed progenitors of various

blood lineages, and in the differentiation and/or proliferation of various cell types. In this aspect of the invention, polynucleotides and polypeptides encoding SEQ ID NO: 293 are introduced into the cells and the cells cultured. These methods may be practiced according to methods well known to the routineer.

Protein of SEQ ID NO:316 (internal designation 188-45-1-0-D9-CS)

The protein of SEQ ID NO:316 is encoded by the cDNA of SEQ ID NO:75. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:316 described throughout the present application also pertain to the polypeptide encoded by a nucleic acid included in clone 188-45-1-0-D9-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:75 described throughout the present application also pertain to the nucleic acid included in clone 188-45-1-0-D9-CS.

The protein of SEQ ID NO:316 is expressed in brain and contains three membrane-spanning segments located between amino acid positions 6 and 26, 73 and 93, or 139 and 159 and a signal peptide comprising the sequence FAAFCYMLSLVLC/AA. Accordingly, one embodiment of the present invention is a polypeptide comprising one or more of the membrane-spanning segments, and/or the signal peptide.

The protein of SEQ ID NO:316 is a member of the cornichon protein family. It has 48% identity with the *Drosophila melanogaster* cornichon protein as well as 67% identity with the Human Cornichon homolog TGAM77 (Genbank accession No. AF104398, the disclosure of which is incorporated herein by reference in its entirety), 67% identity with hCornichon, a bone marrow secreted protein (PCT publication WO/9933979, the disclosure of which is incorporated herein by reference in its entirety), 67% identity with a human secreted protein encoded by gene 24 (PCT publication WO/9910363, the disclosure of which is incorporated herein by reference in its entirety) and 67% identity with the protein product of the mouse *cnih* gene. However, this protein has higher homology, 81% identity, to the mouse cornichon-like protein (Genbank accession No. AB006191, the disclosure of which is incorporated herein by reference in its entirety), which is the product of the mouse *cnil* gene. Finally, the protein of SEQ ID NO:316 has a high level of identity with human secreted protein encoded by gene 95 (GSP:Y76218, PCT publication WO/9958660, the disclosure of which is incorporated herein by reference in its entirety) and is likely a polymorphic varient of gene 95. The high degree of sequence conservation between the members of this family indicates that they are under strong selective pressure and are likely involved in important cellular functions.

The *Drosophila cornichon* (*cni*) gene product is involved in signaling processes necessary for both anterior-posterior and dorsal-ventral pattern formation during *Drosophila* embyrogenesis (Cell, 1995, 81:967-978). Mutations in *cornichon* prevent the formation of a correctly polarized microtubule cytoskeleton in the oocyte. *Cni* signaling functions in concert with two other proteins.

Gurken, which is a protein secreted from the oocyte containing a single epidermal growth factor (EGF) motif most similar in structure to vertebrate TGFα, is considered to be the ligand of the Drosophila epidermal growth factor receptor (DER) homolog torpedo, which is expressed by the follicular epithelium. The function of all three genes in an EGF-like signaling pathway appears to direct the formation of a correctly polarized microtubule cytoskeleton, which is thought to be the basis for the correct spatial localization of other signaling molecules essential for oocyte polarization, asymmetric movement of the nucleus, and embryo differentiation. TGAM77, one of the human homologs of cornichon, is differently expressed in alloactivated T-cells (Bioch. Biophys. Acta 1999, 1449:203-210, the disclosure of which is incorporated herein by reference in its entirety). Since there is a well-known involvement of the microtubule cytoskeleton in spatial polarization of signaling events in T-cell activation, it is thought that TGAM77 may function in a protein-tyrosine kinase pathway required for the vectorial localization of signaling molecules in T-cell activation.

The protein of SEQ ID NO:316 is found in brain tissue, and gene 95 (GSP:Y76218, PCT publication WO/9958660, the disclosure of which is incorporated herein by reference) is expressed in infant brain tissue, endometrial tumor tissue and fontal cortex tissue. ESTs matching this gene are also found in lung tissue, germ cell tumors and skin melanomas. This is similar to the expression pattern of the murine *cnil* gene, which is found in 6.5-day whole embryos, 11.5-day limb bud, 13.5-day whole embryo, adult lung and brain (Dev. Genes Evol., 1999, 209:120-125, the disclosure of which is incorporated herein by reference in its entirety).

Polynucleotides encoding the protein of SEQ ID NO:316 or fragments thereof and polypeptides comprising the protein of SEQ ID NO:316 or fragments thereof are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endometrial tumor, and neural and developmental diseases and/ or disorders. Similarly, the protein of SEQ ID NO:316 or fragments thereof and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural and reproductive organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, reproductive, cancerous and wounded tissues) or bodily fluids (e.g. lymph, serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in infant brain tissue and adult brain tissue, as well as the homology to cornichon proteins, indicates that polynucleotides encoding the protein of SEQ ID NO:316 or fragments thereof and polypeptides comprising the protein of SEQ ID NO:316 or fragments thereof are useful for detecting and/or treating neural and developmental disorders. The tissue distribution

indicates that these polynucleotides and polypeptides are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, Psychoses, autism, and altered behaviors, including disorders in feeding, sleep platterns, balance, and perception. In addition, the gene or gene product may also play a role in treatment and/or detection of developmental disorders associated with the developing embyo, or sexually-linked disorders,

Elevated expression of the protein of SEQ ID NO:316 within the brain suggests that it may be involved in neuronal survival, synapse formation, conductance, neural differentiation, etc. Such involvment may impact many processes, such as learing and cognition. Alternatively, the tissue distribution in endometiral tumor tissue, germ cell tumors and skin melanomas indicates that the translation product of this gene is useful for the detection and/or treatment of endometrial tumors and/or reproductive disorders, as well as tumors of other tissues where expression of this gene has been observed. Furthermore, the protein of SEQ ID NO:316 or fragments thereof may also be used to determine biological activity, to raise antibodies, as a tissue marker, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. The protein of SEQ ID NO:316 or fragments thereof, as well as, antibodies directed against the protein may be used as tumor marker and/or immunotherapy targets for the above listed tissues.

The gene encoding the protein of SEQ ID NO:316 is thought to reside on chromosome 11. Accordingly, polynucleotides encoding the protein of SEQ ID NO:316 or fragments thereof are useful as a marker in linkage analysis for chromosome 11.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:316 , fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition in an individual. For example, the condition may be an abnormality in development, a signaling pathway, microtubule construction, neuronal survival, synapse formation, conductance, neuarl differentiation, or it may be cancer or an abnormality in any of the functions listed above. In such embodiments, the protein of SEQ ID NO:316, or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ ID NO:316. The protein of SEQ ID NO:316 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:316 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:316 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:316 or a cell or preparation containing the protein of SEQ ID NO:316 with a test agent and assaying whether the

test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:316 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:316 may be identified by contacting the protein of SEQ ID NO:316 or a cell or preparation containing the protein of SEQ ID NO:316 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify the source of a sample as, for example, brain, or to distinguish between two or more possible sources of a sample on the basis of the level of the protein of SEQ ID NO:316 in the sample. For example, the protein of SEQ ID NO:316 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from brain or tissues other than brain to determine whether the test sample is from brain. Alternatively, the level of the protein of SEQ ID NO:316 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:316 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other technques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from brain or tissues other than brain to determine whether the test sample is from brain.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:316, including using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:316 or a fragment thereof may be fixed to a solid support, such as a chromatograpy matrix. A preparation containing cells expressing the protein of SEQ ID NO:316 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:316 or a fragment thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:316. In such techniques, the level of the protein of SEQ ID NO:316 in an ill individual is measured using techniques such as those described herein. The level of the protein of SEQ ID NO:316 in the ill individual is compared to the level in normal individuals to determine whether the individual has a level of the protein of SEQ ID NO:316 which is associated with disease.

Protein of SEQ ID NO:255 (106-037-1-0-E9-CS.cor)

The protein of SEQ ID NO:255, encoded by the cDNA of SEQ ID NO:14, is strongly expressed in the liver and testis and shows extensive homology to human lactate dehydrogenase-A protein (LDH-A or M chain) (Chung F.Z. et al., Biochem. J. 231:537-541(1985); SwissProt accession number P00338). The protein of SEQ ID NO:255 is also homologous to lactate dehydrogenase A from many vertebrates. The 381-amino-acid-long protein of SEQ ID NO:255 displays a Prosite motif corresponding to lactate dehydrogenase from positions 71 to 380. In addition, the active site LGEHGDS, where H is the active site residue, is present in the protein of the invention (positions 239 to 245). The protein of the invention also contains an additional 50 N-terminal amino acids not found in other lactate dehydrogenase A proteins. This N-termimal extension contains a signal peptide (cleavage site at position 34 of the protein of invention) that may allow the export of the protein to the extracellular domain or define a particular subcellular localization. Alternatively, the initiation start codon could be at position 26 or 50 of the protein of SEQ ID NO:255.

Lactate dehydrogenase (LDH) is an enzyme which dehydrogenates lactic acid into pyruvic acid in conjunction with the hydrogen acceptor NAD+, and which exists in a wide variety of animal tissues and microorganisms as an enzyme serving to produce lactic acid from pyruvic acid in the glycolytic pathway (Abad-Zapatero C. et al. J. Mol. Biol. 198:445-467(1987)). It is known that in vertebrates there are three isozymes of LDH: the M form (LDH-A), found predominantly in muscle tissues; the H form (LDH-B), found in heart muscle, and the X form (LDH-C), found only in the spermatozoa of mammals and birds. In birds and crocodilian eye lenses, LDH-B serves as a structural protein and is known as epsilon-crystallin (Hendriks W. et al. Proc. Natl. Acad. Sci. U.S.A. 85:7114-7118(1988)).

LDH has been used extensively in the field of clinical test reagents for a number of purposes. For example, it has been used as a coupling enzyme to determine the enzymatic activity of various amino-transferases, such as alanine aminotransferase (ALT), which is ultimately detected by UV spectrometry of the produced pyruvic acid. This use of LDH has been widely adopted as a clinical test, because amino-transferases are enzymes which

show high activity in liver, heart, kidney, etc. and show remarkable increases in serum in association with various diseases. LDH has also been used as a coupling enzyme to help determine the level of substrates such as urea, as the enzyme promotes the conversion of such substances into pyruvic acid which can be detected by UV spectrometry.

Lactate dehydrogenase is also a widely used marker for heart disease and other conditions. For example, levels of LD-1 are elevated in the presence of myocardial infarction and in other conditions such as leukemia. Levels of lactate dehydrogenase start to increase 24 to 48 hours after occlusion of the coronary artery, peak in 3 to 6 days, and return to normal in 8 to 14 days. In addition, levels of LD-1 are elevated 10 to 12 hours after the acute myocardial infarction, peak in 2 to 3 days, and return to normal in approximately 7 to 10 days. Thus, measurement of the level of lactate dehydrogenase allows a prolonged retrospective diagnosis of myocardial infarction. Further, while the amount of LD-2 in the blood is usually higher than the amount of LD-1, patients with acute myocardial infarction have more LD-1 than LD-2. This "flipped ratio" usually returns to normal in 7 to 10 days. An elevated level of LD-1 with a flipped ratio has a sensitivity and specificity of approximately 75% to 90% for detection of acute myocardial infarction.

Elevated LDH levels have also been used as a prognostic indicator for cancers such as small cell lung carcinoma. Specifically, elevated levels of LDH indicate a poor prognosis for such diseases (Kawahara, et al., (1997) Jpn J Clin Oncol. 1997 Jun;27(3):158-65).

LDH expression in cells has also been shown to be induced by interleukin-1 alpha, a major cytokine associated with, e.g., inflammation (Nehar et al. (1998) Biol Reprod Dec;59(6):1425-32).

Islet beta-cells express low levels of lactate dehydrogenase and have high glycerol phosphate dehydrogenase activity. The effects on glucose metabolism and insulin secretion of acute overexpression of the skeletal muscle isoform of lactate dehydrogenase (LDH)-A in these cells have been studied by Ainscow EK et al. (Diabetes 2000 Jul;49(7):1149). The results of these studies have shown that overexpression of LDH activity interferes with normal glucose metabolism and insulin secretion in islet beta cells, and it may therefore be directly responsible for insulin secretory defects in some forms of type 2 diabetes. These results also reinforce the view that glucose-derived pyruvate metabolism in the mitochondria is critical for glucose-stimulated insulin secretion in beta cells. Other data show that an overexpression of lactate dehydrogenase A attenuates glucose-induced insulin secretion in stable MIN-6 beta-cell lines, which normally express low levels of L-lactate

dehydrogenase (Zhao C, Rutter GA FEBS Lett. 1998 Jul 3;430(3):213-6). Low LDH activity thus appears to be important in beta-cell glucose sensing.

Analysis of the LDH isoenzyme pattern in CSF fluid has also been shown to be helpful in the evaluation of CNS involvement in patients with hematologic malignancies (Lossos IS, et al. Cancer. 2000 Apr 1;88(7):1599-604).

It is believed that the protein of SEQ ID NO:255 is a lactate dehydrogenase protein, most likely of the LDH-A or M subtype. The activity of the present protein can be assessed using any standard method for detecting lactate dehydrogenase enzyme activity, including those involving the UV detection of pyruvate, a product of LDH-catalyzed enzymatic reactions.

In one embodiment, the polypeptides and polynucleotides of the invention are used to detect testis and liver tissue, as well as cells derived from these tissues. For example, nucleic acids and proteins of the invention can be labeled isotopically or chemically, using methods known to those skilled in the art, and used as probes in northern blots, far-western blots and *in situ* hybridization experiments. An ability to detect specific cell types is useful, e.g. for the determination of the history of tumor cells, as well as for the identification of cells and tissues for histological studies.

In another embodiment, the present protein can be used in any of a variety of clinical assays involving LDH enzymes. For example, the protein can be used as a coupling enzyme to determine the enzymatic activity of various amino-transferases, such as alanine aminotransferase (ALT), as detected by UV spectrometry of the produced pyruvic acid. Such assays have significant clinical utility, as amino-transferases are enzymes which show high activity in liver, heart, kidney, etc. and show remarkable increases in serum in association with various diseases. The protein of the invention can also be used as a coupling enzyme to help determine the level of substrates such as urea, as the enzyme promotes the conversion of such substances into pyruvic acid which can be detected by UV spectrometry.

In another embodiment, the present protein can be used to identify ingredients for cosmetic formulations. Specifically, enhancers of lactate dehydrogenase can be included in cosmetic compositions to stimulate keratinocyte proliferation and collagen synthesis in cutaneous tissues. The inhibitors can be combined with other active ingredients such as pyruvic acid, acetic acid, acetoacetic acid, beta-hydroxybutyric acid, Krebs cycle pathway metabolites, aliphatic saturated or unsaturated fatty acids containing from 8 to 26 carbon atoms, omega-hydroxy acids containing from 22 to 34 carbon atoms, glutamic acid, glutamine, valine, alanine, leucine, and mixtures thereof (see, e.g., US Patent 5,853,742, the disclosure of which is hereby incorporated by reference in its entirety).

In another embodiment, the present invention provides methods for treating or preventing cancer, e.g., by inhibiting lactate dehydrogenase activity in cells, preferably specifically the cancer cells, of a patient. The expression or activity of lactate dehydrogenase can be inhibited using any of a large number of agents, including, but not limited to, antibodies, antisense molecules, ribozymes,

and heterologous molecules that inhibit the expression or activity of the lactate dehydrogenase in the cancer cells of the patient. In one embodiment, lactate dehydrogenase that has been obtained from a primate, or anti-lactate dehydrogenase antibodies obtained from a mammal as a result of the parenteral administration of primate lactate dehydrogenase to said mammal, is parenterally administered to human cancer patients. Antibodies derived from the protein of the invention or part thereof can also be used to inhibit cancer cell development as described in US Patent No. 4,620,972.

Analysis of the LDH isoenzyme pattern in CSF fluid has been shown to be helpful in the evaluation of CNS involvement in patients with hematologic malignancies (Lossos IS, et al. Cancer. 2000 Apr 1;88(7):1599-604). Thus, in another embodiment, the protein of SEQ ID NO:255 can be used to develop assays to monitor the LDH isoenzyme activity in CSF fluid, thereby improving the sensitivity of CSF cytology. This assay may be derived, e.g., from the methods described by Short S. et al. (J Biol Chem. 2000 Apr 28;275(17):12963-9).

In another embodiment, the protein of SEQ ID NO:255 is used to detect and/or treat insulin secretory defects in some forms of type 2 diabetes. For example, various evidence indicates that LDH overexpression may be involved in certain types of diabetes. Therefore, the detection of an elevated level of LDH in a patient, e.g. in pancreatic islet cells of a patient, can be used as an indication that the patient has diabetes, or is at risk of developing diabetes. Similarly, methods of inhibiting the expression or activity of LDH in those cells, e.g. using antibodies, antisense sequences, or heterologous compounds that inhibit the expression or activity of LDH, can be used to treat or prevent diabetes.

In another embodiment, the protein of the invention can be used to eliminate endogenous pyruvic acid in cells in vitro or in vivo.

In another embodiment, the expression of the present protein is used as a marker for interleukin 1, e.g. IL-1 alpha, activity in cells or in a patient. Specifically, as it has been shown that LDH expression is induced by IL-1 alpha, then the expression, or elevated expression, of the present protein can be used as a marker for the action of IL-1 on the cell. As IL-1 has been implicated in a number of physiological processes, including inflammation and more specifically in deleterious processes such as arthritis and autoimmune disorders, the present protein can serve as a marker for the presence of such disorders, or for a predisposition for the disorders.

In another embodiment, the present protein is used to detect heart disease and other diseases in patients. For example, levels of LDH are known to rise following myocardial infarction and other heart ailments. Accordingly, the detection of an elevated level of the protein of the invention, alone or in view of the levels of other proteins such as other LDH isozymes, can be used as an indicator of a heart attack or other diseases, including leukemia. The levels of LDH can be assessed in any tissue or biological sample, including, but not limited to, serum, and can be detecting using any standard method, including, but not limited to, immunoassays and assays for LDH enzyme activity.

In another embodiment, the present protein is used to determine a prognosis for any of a number of diseases, including cancers such as small cell lung carcinoma. For example, the level of the present protein is detected in the serum of a patient suffering from cancer, wherein the detection of a decreased level of expression or activity of the protein indicates a worse prognosis for the patient compared to the prognosis in a patient with a normal level of the protein activity or expression.

Proteins of SEQ ID NOs:243, 253 (internal designation numbers 105-016-1-0-D3-CS and 105-095-2-0-G11-CS)

The 331-amino-acid- long protein of SEQ ID NO:243, encoded by the cDNA of SEQ ID NO:2, is found in prostate and in fetal brain and is homologous to a secreted human protein (Genseq accession number Y59685). In addition, this protein is highly homologous to the the putative glycerophosphodiester phosphodiesterase (GP-PDE) MIR16 (Membrane Interacting protein of RGS16) protein (SPTREMBLNEW SPTREMBL SWISSPROT accession number AAF65234) encoded by the cDNA of GENPEPT GENPEPTNEW accession number AF212862; in fact, the protein of the invention is a likely variant of the MIR16 protein. Furthermore, a BLAST search with the amino acid sequence of SEQ ID NO:243 indicates that the protein of the invention is homologous to GP-PDEs of *E.coli* (SWISSPROT accession numbers P09394 and P10908) and *Haemophilus influenzae* (SWISSPROT accession number Q06282). The protein of SEQ ID NO:243 displays 2 candidate membrane-spanning segments, from amino acids 7 to 27 and 258 to 278, and a putative signal peptide from amino acids 19 to 24. Finally, the protein of the invention has two putative *N*-glycosylation sites: asparagine residues at positions 168 and 198 (Zheng *et al.*, Proc. Natl. Acad. Sci. 97 :3999-4004 (2000)).

The cDNA of SEQ ID NO:2 differs from the cDNA of GENPEPT GENPEPTNEW accession number AF212862 by its extended 5' and 3' termini, and from the cDNA of SEQ ID NO:12 by polymorphisms and alternate splicings.

The MIR16 (Membrane Interacting protein of RGS16) protein, which is homologous to the protein of the invention, was identified in a yeast two-hybrid screen of a pituitary cell cDNA library using the RGS16 (Regulator of G protein Signaling) protein as bait (Zheng *et al.*, Proc. Natl. Acad. Sci. 97:3999-4004 (1999)). and Sasaki, J. Bacteriol. 175:4569-4571 (1993); Zheng *et al.*, ibid.). Remarkably, the GP-PDE from *Haemophilus influenzae* (also called protein D) which is 67% identical to the periplasmic GP-PDE of *E.coli*, presents affinity for human immunoglobin D (Janson et al., Infect. Immun. 62:4848-854 (1994)).

From sequence alignments, it can be seen that the N-terminal region of MIR16 (amino acids 70-150), immediately after the putative signal peptide, is highly conserved (40-61% similarity), suggesting that it may contain residues critical for catalytic activity, i.e., the catalytic site. GP-PDEs hydrolyze deacetylated phospholipid GPs, such as glycerophosphocholine (GPC)

and glycerophosphoethanolamine, to sn-glycerol-3-phosphate (G3P) and the corresponding alcohols (Zheng *et al.*, ibid.). The putative enzymatic activity of MIR16 and its interaction with RGS16 suggest that it may play important roles in lipid metabolism and in G protein signaling. As shown in northern blot experiments, the MIR16 mRNA is highly transcribed in heart, liver, kidney, testis and brain. The observed expression of MIR16 in the brain is consistent with the above-described expression of the protein of the invention in the fetal brain.

It is believed that the proteins of SEQ ID NOs:243 and 253 or part thereof are members of the glycerophosphodiester phosphodiesterase protein family, interact with the RGS16 protein and, as such, play important roles in both lipid metabolism and in G protein signaling. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:243 from positions 7 to 27, 19 to 24 and 258 to 278. Other preferred polypeptides of the invention are fragments of SEQ ID NO:243 or 253 having any of the biological activities described herein. Additional preferred polypeptides are those that comprise asparagine residues at positions 168 and/or 198.

The invention first relates to methods and compositions using cDNAs of SEQ ID NO:2 or 12 or part thereof, and proteins of the invention SEQ ID NO:243 or 253 or part thereof to identify specific cell types, preferably from prostate or fetal brain. For example, nucleic acids and proteins of the invention are labeled isotopically or chemically following methods known to those skilled in the art, and further used as probes in northern blots, far-western blots and *in situ* hybridization detection experiments. An ability to detect specific cell types is useful, e.g. for the determination of the history of tumor cells, as well as for the identification of cells and tissues for histological studies.

Any of a number of in vitro assays can be used to detect SEQ ID NO:243 or 253 protein activity, for example for *in vitro* screening of modulators of protein activity. Preferably cDNA encoding the protein of the invention is cloned in a prokaryotic expression vector, according to methods known to those skilled in the art. Briefly, the GP-PDE activity of the recombinant protein is analyzed by a coupled spectrophotometric assay as described by Larson and collaborators and adapted by Cameron and collaborators (Larson et al., J. Biol. Chem. 258 :5426-5432 (1983); Cameron et al., Infect. Immun. 66 :5763-5770 (1998)). Such enzymatic activity may be measured *in vitro* in the presence of modulating drugs.

Another embodiment of the present invention relates to methods of using the protein of the invention or part thereof to purify or specifically bind to human immunoglobin D. Several immunoglobin (Ig) binding bacterial cell wall proteins have been isolated and/or cloned during the last two decades. The best characterized of these are protein A of Staphylococcus aureus (which binds to human IgG subclasses 1, 2 and 4, IgG of several mammalians species, and in some instances human Ig of classes A, M, E), and protein G of group G beta-hemolytic streptococci (which binds to all human IgG subclasses and which also displays a wider binding spectrum for

animal IgG than protein A). IgD binds to neither protein A nor protein G. Consequently, it is of great interest to identify new proteins capable of binding IgD, thereby allowing its separation and purification. In addition, IgD binding proteins can also be used in immunoprecipitation procedures with IgD, as are routinely performed with proteins A and G in the case of IgG. The binding and purification of IgD using the protein of the invention can be accomplished in any of a number of ways, for example by generating a fusion protein or polypeptide in which the protein of the invention or part thereof, is combined with another protein by the use of a recombinant DNA molecule. The resulting fusion product including the protein of the invention or part thereof is then covalently, or by any other means, bound to a protein, carbohydrate or matrix (such as gold, "Sephadex" particles, polymeric surfaces). Such a complex is very useful for IgDs immobilization and consecutive immunoprecipitations in batch. Similar assays for binding of protein D (GP-PDE) of *Haemophilus influenzae* and IgD are described in the US Patent No. 6,025,484.

Another embodiment of the invention relates to compositions and methods using the protein of the invention, or part thereof, as GP-PDE enzymes to hydrolyze deacylated phospholipids (GPs), such as glycerophosphocholine (GPC) and glycerophosphoethanolamine, to sn-glycerol-3-phosphate (G3P) and the corresponding alcohols. First, this enzymatic activity, which belongs to the class of specific phospholipase D, makes the protein of the invention very useful to study biological membranes and their phospholipidic components. Moreover, as glycerophospholipids are major components of the lipidic bilayer, elimination of their hydrophilic moiety using the GP-PDE activity of the protein of the invention would likely modify the structure and consequently the permeability of eukaryotic cell membranes. Such modifications could improve the transfection efficiency of eukaryotic cells, in vitro or in vivo. Typically, in such embodiments the purified protein of SEQ ID NOs:243 or 253 is administrated to cells; purified proteins of the invention can be obtained in any of a number ways, for example by inserting the cDNA encoding the proteins into a prokaryotic expression vector using any technique known to those skilled in the art. The recombinant protein produced and purified in the prokaryotic system is then added to an *in vitro* culture of eukaryote cells before or during transfection. The recombinant protein of the invention can also be used to increase the efficiency of cell transfection *in vivo,* most notably in the case of gene therapy. For example, tumoral masses are very often resistant to transfection, and the protein of the invention would likely provide an effective way to facilitate the introduction of cytotoxic genes (such as pro-apoptotic genes) or antitumoral drugs in solid tumors.

Still another embodiment of the protein of the invention relates to methods and compositions to diagnose, treat, and prevent disorders associated with excess glutamate signaling in the brain. As described above, the MIR16 protein interacts physically with the RGS16 protein (Regulator of G protein Signaling 16). Receptors of many hormones use heterotrimeric G proteins for signal transduction after ligand binding (for a review, see Neer, Cell 80 :249-257 (1995)). Among these receptors are metabotropic glutamate receptors (mGluRs). These receptors, which are

expressed in the brain, like the protein of the invention, are a novel family of cloned G-protein-coupled receptors (Schoepp and Conn, Trends Pharmacol. Sci. 14:13-20 (1993)). Endogenous glutamate, by activating the mGluR1 receptor (and also NMDA and AMPA receptors), may contribute to the brain damage occurring acutely after epilepsy, cerebral ischemia or traumatic brain injury. It may also contribute to chronic neurodegeneration in such disorders as amyotrophic lateral sclerosis and Huntington's chorea (Meldrum, J. Nutr. 130(4S Suppl):1007S-1015S (2000)).

The invention thus relates to methods and compositions using cDNAs of SEQ ID NO:2 or 12 or part thereof, and proteins of SEQ ID NO:243 or 253 or part thereof, to diagnose, treat, or prevent disorders associated with excess glutamate signaling in the brain. Specifically, the level of activity or expression of the proteins can be correlated with the level of glutamate signaling, or with the glutamate-signaling associated brain damage involved in epilepsy, cerebral ischemia, traumatic brain damage, ALS, or Huntington's chorea, or with any other G-protein associated physiological process or disease or condition. For situations where the level of the expression or activity of the protein is positively correlated with such signaling or with the presence of a disease or condition, the signaling, disease or condition can be detected using any of a number of tools for detecting protein expression or activity, including northern blots, far-western blots and *in situ* hybridization experiments, where an elevated level of the protein, protein activity, or nucleic acid of the invention indicates the presence of the disease, condition, or signaling process. Further, such diseases or conditions can be treated or prevented, or such signaling pathways can be inhibited, using compounds that inhibit the expression or activity of the protein, such as antibodies, antisense molecules, ribozymes, dominant negative forms of the protein, or any heterologous molecule that inhibits protein activity or expression. Alternatively, where the expression or activity of the protein of the invention is negatively associated with the signaling pathway, disease or condition, a detection of a decreased level of expression or activity of the protein can be used to indicate the presence of the disease, condition, or pathway. Further, in such cases, the disease or condition can be treated or prevented, or the pathway be inhibited, using any compound that increases the activity or level of the protein, such as nucleic acids encoding the protein, the protein itself, or heterologous compounds that cause an increase in the level of protein expression or activity.

Protein of SEQ ID NO:386 (internal designation 105-037-4-O-H12-CS)

The protein of SEQ ID NO:386, encoded by the cDNA of SEQ ID NO:145, is strongly expressed in the fetal brain and uterus. The 207-amino-acid-long protein of SEQ ID NO:386 displays pfam SPRY domains from positions 85 to 205.

SPRY domains have been found in a number of proteins involved in multiple cellular and developmental processes. For example, the Midline-1/FXY family of proteins has been shown to associate with microtubules, and has been implicated in human diseases, such as Opitz Syndrome, a congenital disorder characterized by multiple developmental abnormalities (see, e.g., Cainarca, et

al., (1999) Hum Mol Genet 8(8):1387-96). In addition, the cytoplasmic Marenostrin/Pyrin protein has been demonstrated to be the cause of Familial Mediterranean fever, an autosomal recessive disorder characterized by fever and serositis (Nat Genet 1997 Sep;17(1):25-31). Other SPRY proteins include SplA, a serine protease from Staphylococcus aureus, and butyrophilin, a major milk protein. Another family of proteins known to contain the SPRY domain are the Ryanodine receptors (RyRs).

Ryanodine receptors play an important role in Ca2+ signaling in muscle and non muscle cells by releasing Ca2+ from intracellular stores. For example, these receptors are centrally important in excitation–contraction (e-c) coupling, which occurs at specialized regions where the sarcoplasmic reticulum (SR), containing the ryanodin receptors, and the plasma membrane/transverse-tubule system form junctions. RyRs are also thought to play some role in maintaining the structural integrity of the SR·T-tubule junctions. RyR is apparently unable to carry out the requisite functions associated with e–c coupling by itself, however, because it forms interactions with other macromolecules at the triad junction. For example, two small proteins, calmodulin and FKBP12, are believed to modulate RyR at the triad junction.

It is believed that mammalian tissues express three different RyR isoforms, comprising four 560-kDa (RyR polypeptide) and four 12-kDa (FK506 binding protein) subunits. It is believed that these large protein complexes conduct monovalent and divalent cations and are capable of multiple interactions with other molecules. The subunits of the protein complexes include small diffusible endogenous effector molecules including Ca2+, Mg2+, adenine nucleotides, sufhydryl modifying reagents (glutathione, NO, and NO adducts) and lipid intermediates, and proteins such as protein kinases and phosphatases, calmodulin, immunophilins (FK506 binding proteins), and in skeletal muscle the dihydropyridine receptor. The RyR from skeletal muscle is the major calcium release channel for that tissue, and the most intensively studied of the three genetic isoforms detected thus far in mammalian species. The other two RyR isoforms are often referred to as the 'heart' and 'brain' forms, but the actual cell and tissue distribution of the isoforms is complex.

Because of their multiple ligand interactions, ryanodin receptors constitute an important, potentially rich pharmacological target for controlling cellular functions. Ca2+ release channel activity is modulated by many endogenous effectors, including Ca2+, ATP, Mg2+, and calmodulin. In addition, many exogenous effectors, including caffeine, local anesthesics, and polyamines, also modify channel activity. For example, tetracaine, procaine, benzocaine, and lidocaine inhibit Ca2+ release from the SR. They appear to interact with a specific site(s) located on the RYR, affecting both ryanodin-binding and single channel activities (Shoshan-Barmatz et al. 1993; J. Membr. Biol.; 133; 171-181).

The importance of intracellular calcium as a second messenger in cellular signal transduction processes is well established. Alterations in intracellular Ca2+ homeostasis have profound effects on many cell functions, including secretion, contraction-relaxation, motility,

metabolism, protein synthesis, modification and folding, gene expression, cell-cycle progression and apoptosis. A major source of cytoplasmic calcium is from intracellular storehouses located in the endoplasmic reticulum, or in muscle, within the sarcoplasmic reticulum (SR).

Given that cellular Ca2+ handling is an important factor in the control of neuronal metabolism and electrical activity, abnormalities of intracellular Ca2+ channels might be expected to contribute to some forms of epilepsy or to anoxic brain damage following an episode of cerebral ischemia. Cell loss is said to be a characteristic feature of degenerative brain disorders, including Alzheimer's disease. It is well established that neuronal cell death may be secondary to an abnormal elevation of cytoplasmic Ca2+, particulary that associated with activation of excitatory glutamate receptors (e.g., in epilepsy). This strongly suggests that the release of stored Ca2+ contributes to nerve cell damage and cell death in various circumstances.

It is believed that the protein of SEQ ID NO:386 is functionally related to other SPRY-containing proteins, such as the ryanodine receptors, Marenostrin/Pyrin, SplA, Midline-1/FXY, and butyrophilin. Accordingly, it is thus believed that the present protein is associated with the release of Ca2+ from intracellular Ca2+-storing organelles, like the endoplasmic reticulum and, in muscle, the sarcoplasmic reticulum (SR), as well as being involved in microtubule binding. Preferred polypeptides of the invention are any fragments of SEQ ID NO:386 having any of the biological activities described herein.

In one embodiment, the present protein and nucleic acids can be used to specifically detect cells of the fetal brain and uterus, as the protein is overexpressed in these tissues. For example, the protein of the invention or part thereof may be used to synthesize specific antibodies using any technique known to those skilled in the art. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, such as in forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, etc., or to differentiate different tissue types in a tissue cross-section using immunochemistry. The protein can also be used to specifically label microtubules in cells.

In another embodiment, the protein of the invention or part thereof may be used in regulating intracellular Ca2+ levels. As alterations in intracellular Ca2+ homeostasis have profound effects on many cell functions, including secretion, contraction-relaxation, motility, metabolism, protein synthesis, modification and folding, gene expression, cell-cycle progression and apoptosis, the ability to modulate intracellular Ca2+ levels provides a tool to alter any of these cellular functions, in vitro or in vivo. Such an ability has wide utility for a large number of applications, for example to manipulate the behavior (e.g. growth rate, secretion, survival, etc.) of cells grown in vitro, as well as to treat, prevent, or diagnose any of a number of diseases associated with altered Ca2+ signaling in vivo. The activity or expression of the protein of the invention can be modulated in any of a large number of ways, for example by administering to cells or to a patient the protein itself, a polynucleotide encoding the protein, antibodies, antisense sequences, dominant negative

forms of the protein, compounds that alter the expression or activity of the protein, etc. The effect of any such agent on calcium flux in cells can be detected using standard methods, including by studying the permeation of Ca2+ release through endoplasmic reticulum (ER) and sarcoplasmic reticulum (SR) channels using tracers, light scattering and fluorescence quenching, and channel reconstitution in planar bilayer. In addition, targeted recombinant photoproteins can provide direct measurements of organellar Ca2+ (Montero et al.; 1995; EMBO J.; 14, 5467-5475).

The invention further relates to methods and compositions using the protein of the invention or part thereof to diagnose, prevent and/or treat several disorders in which the activity or recognition of ryanodin receptors, is impaired or excessive. These disorders include, but are not limited to, neurodegenerative diseases, cardiovascular disorders, severe myasthenia, malignant hyperthermia, epilepsy, and central core disease. For example, in patients with severe myasthenia, the level of anti-RyR antibodies has been directly related to the severity of the disease (Skeie et al., 1996: Eur. J. Neurol. 3; 136-140). There is also some evidence to suggest that RyR abnormalities are a primary cause of many types of cardiac disease. In addition, the protein of the invention can be used to diagnose other diseases associated with SPRY-protein dysfunction, such as Familial Mediterranean fever and Opitz syndrome. Finally, as SPRY containing proteins have been implicated in embryonic development (e.g. the Midline1 protein), the protein and nucleic acids of the invention can be used to detect developmental disorders, as the detection of a mutation in the gene encoding SEQ ID NO:386, or a detection of abnormal gene expression in a fetus, can be used to indicate the presence of a developmental abnormality. For example, as the protein of SEQ ID NO:386 is strongly expressed in the fetal brain, it is likely that the protein plays a role in the normal development of the brain in utero.

The present invention also relates to diagnostic assays for detecting altered levels of the protein of SEQ ID NO:386 in various tissues, as over-expression of the protein compared to normal control tissue samples can indicate the presence of certain disease conditions such as neurodegenerative disorders, cardiovascular disorders, svere myasthenia, malignant hyperthermia, epilepsy, and central core disease. Assays used to detect levels of the polypeptide of the present invention in a sample derived from a host are well-known to those of skill in the art and include radioimmunoassays competitive-binding assays, Western Blot analysis and ELISA assays.

Proteins SEQ ID NOs:283 and 286 (internal designations 174-38-1-0B6-CS_LA and 174-41-1-0-A6-CS LA)

The protein of SEQ ID NO:283, encoded by the cDNA of SEQ ID NO:42, is overexpressed in salivary glands and to a lesser extent in bone marrow, and shows homology over the C-terminal length to the immunoglobin (Ig) protein superfamily, which is conserved among eukaryotes (including rabbit, rodents and human). In particular, the 468-amino-acid-long protein of the invention, which is similar in size to the constant chain of Ig related proteins, displays two pfam

conserved immunoglobulin domains, from position 205 to 285 and from position 318 to 384, which are known to be involved in the basic structure of the light and heavy constant chains of immunoglobins. It is known (Orr H.T., Nature 282:266-270(1979)) that the Ig constant chain domains and a single extracellular domain in each type of MHC chain are closely related, sharing over one hundred amino-acids of homology. All members of the Ig related superfamily, including the MHC class I alpha chain and beta-2-microglobulin, as well as the MHC class II alpha and beta chains, display the prosite conserved characteristic pattern around the C-terminal cysteine ([FY]-x-C-x-[VA]-x-H). This cysteine is involved in the disulfide bond between the light and heavy chains, and is also found in the protein of the invention (position 380 to 386). The protein of the invention also exhibits an emotif Ig and Major Histocompatibility Complex protein signature from positions 319 to 336. In addition, the protein of the invention displays homology with tapasin (GeneBank No. AF009510), a chaperone-like protein closely associated with TAP-binding proteins, which is well conserved among eukaryotes (chicken, rodents and human). Tapasin has been shown to increase the efficiency of antigen processing and presentation by mediating the association of MHC complex proteins with TAP proteins to the endoplasmic reticulum and to the cell surface during immune response (for review see Abele, R. and Tampé, R., Bioch. et Biophysica Acta, 1999). In addition, the protein of the invention displays two transmembrane domains from positions 199 to 219 and from positions 406 to 426 , a hydrophobic profile similar in amino acid position to the hydrophobic stretch of amino acids of human and mouse tapasin (Suling L., J. Biol. Chem., 274:8649-8654, 1999), and a secreted signal peptide from position 9 to 23. Both signatures are largely present in Ig related proteins such as secreted antibodies or antigen presenting proteins. The invention also encompasses a variant (SEQ ID NO:286) of SEQ ID NO:283, encoded by the cDNA of SEQ ID NO:45. The protein of SEQ ID No:286 is a 442-amino-acid-long protein with a C-terminal shorter end of 26 amino-acids compared to the protein of SEQ ID NO:283. The variant of SEQ ID NO:286, which results from a frameshift (position 1445 in SEQ ID NO:45) in the coding sequence that leads to a stop codon in the corresponding protein, displays characteristics identical to those described above in terms of motifs, Ig signatures, function, and potential uses.

The immunoglobulin (Ig) gene superfamily comprises a large number of cell surface glycoproteins that share sequence homology with the V and C domains of antibody heavy and light chains. These molecules function as receptors for antigens, immunoglobulins and cytokines as well as adhesion molecules, and play important roles in regulating the complex cell interactions that occur within the immune system (A. F. Williams et al., Annu. Rev. Immuno. 6:381-405, 1988, T. Hunkapiller et al., Adv. Immunol. 44:1-63, 1989; for a short review see also Prosite entry PS00290)

The introduction of an antigen into a host initiates a series of events culminating in an immune response. In addition, self-antigens can result in immunological tolerance or activation of an immune response against self-antigens. A major portion of the immune response is regulated by presentation of antigen by major histocompatibility complex molecules. MHC molecules bind to

peptide fragments derived from antigens to form complexes that are recognized by T cell receptors on the surface of T cells, giving rise to the phenomenon of MHC-restricted T cell recognition. The ability of a host to react to a given antigen (responsiveness) is influenced by the spectrum of MHC molecules expressed by the host. Responsiveness correlates with the ability of specific peptide fragments to bind to particular MHC molecules.

There are two types of MHC molecules, class I and class II, each of which comrise two chains. In class I [2], the alpha chain is composed of three extracellular domains, a transmembrane region, and a cytoplasmic tail. The beta chain (beta-2-microglobulin) is composed of a single extracellular domain. In class II [3], both the alpha and the beta chains are composed of two extracellular domains, a transmembrane region and a cytoplasmic tail. MHC class I molecules are expressed on the surface of all cells, and MHC class II molecules are expressed on the surface of antigen presenting cells. MHC class II molecules bind to peptides derived from proteins made outside of an antigen presenting cell. In contrast, MHC class I molecules bind to peptides derived from proteins made inside a cell. In order to present peptide in the context of a class II molecule, an antigen presenting cell phagocytoses an antigen into an intracellular vesicle, in which the antigen is cleaved, bound to an MHC class II molecule, and then returned to the surface of the antigen presenting cell.

Major histocompatibility complex (MHC) class I molecules present antigenic peptides to CD8 T cells (Townsend, A. et al., Nature:340,443-448)). The peptides are generated in the cytosol and then translocated across the membrane of the endoplasmic reticulum by the transporter associated with antigen processing (TAP). TAP is a trimeric complex consisting of TAP1, TAP2, and tapasin (TAP-A). TAP1 and TAP2 are required for the peptide transport. Tapasin mediates the interaction of MHC class I HC-beta-2 microglobulin with TAP, and this interaction is essential for peptide loading onto MHC class I HC-beta-2-microglobulin (Suling et al., J. Biol. Chem., 274:8649-8654). T cell receptors (TCRs) are the second antigen recognition molecules, and recognize antigens that are bound by MHC molecules. Recognition of MHC complexed with peptide (MHC-peptide complex) by TCR can effect the activity of the T cell bearing the TCR. Thus, MHC-peptide complexes are important in the regulation of T cell activity and, thus, in regulating an immune response.

Human cytomegalovirus (HCMV) is a betaherpesvirus which causes clinically serious disease in immunocompromised and immunosuppressed adults, as well as in some infants infected in utero or perinatally (Alford, C. A., and W. J. Britt. 1990. Cytomegalovirus, p. 1981-2010. In D. M. Knipe and B. N. Fields (ed.), Virology, 2nd ed. Raven press, New York). In human cytomegalovirus (HCMV)-infected cells, expression of the cellular major histocompatibility complex (MHC) class I heavy chains is down-regulated, where down-regulation is defined as reduction in either synthesis, stability or surface expression of MHC class I heavy chains. A similar phenomenon has been reported for some other DNA viruses, including adenovirus, murine

cytomegalovirus, and herpes simplex virus (Anderson, M., et al., Cell 43:215-222, 1985; Burgert and Kvist, Cell 41:987-997, 1985; Heise T. M., et al., J. Exp. Med. 187:1037-1046, 1998). In the adenovirus and herpes simplex virus systems, the product of a viral gene which is dispensable for replication in vitro is sufficient to cause down-regulation of MHC class I heavy chains (Anderson, M., et al., 1985, supra). The gene(s) involved in class I heavy chain down-regulation by murine cytomegalovirus have not yet been identified.

It is believed that the proteins of SEQ ID NOs:283 and 286 are members of the immunoglobulin superfamily and, as such, play a role in the immune response, cellular proteolysis, cell proliferation and differentiation, pathogen recognition, apoptosis, and other processes associated with the Ig superfamily. In addition, the proteins of the invention are thought to be tightly linked to the antigen processing and presentation system in the context of peptide assembly and translocation of foreign peptides across endoplasmic reticulum and cell surface membranes as new chaperonin-like proteins associated with MHC I and TAP proteins. The weak homology (30%) with the TAP protein family is thought to indicate the specificity of the interactions of the proteins of the invention with MHC proteins and/or TAP-related proteins, as described by Suling et al., supra.

Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:283 from position 9 to 23, 199 to 219, 205 to 285, 318 to 384, 319 to 336, 380 to 386 and from 406 to 426. Other preferred polypeptides of the invention are fragments of SEQ ID NO:283 having any of the biological activities described herein.

In one embodiment, the invention relates to methods and compositions for using the protein of the invention or part thereof as a marker protein to selectively identify tissues, such as salivary glands and bone marrow tissues, which strongly express the protein of the invention. For example, the protein of the invention or part thereof may be used to synthesize specific antibodies using any techniques known to those skilled in the art including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry.

In another embodiment, the invention relates to methods for using the protein of the invention to visualize proteins and peptides involved in antigen recognition system within cells by virtue of their physical interaction with the proteins of the invention. For example, the protein may be used to detect the presence and/or the localization of MHC peptides and TAP- like proteins in a cell. The protein of the invention, and hence any interacting proteins, can be labeled using any of a number of methods, including by binding with specific antibodies or by creating a fusion protein comprising the protein of the invention as well as a readily detectable moiety, such as an epitope tag, biotin, or green fluorescent protein.

In another embodiment, polynucleotide or polypeptide sequences of the invention or part thereof may be used for the diagnosis of a disorder associated with a loss of regulation of the expression of the protein of the invention, preferably, but not limited to, deficiencies of the MHC protein system. Examples of such disorders include, but are not limited to, acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease; a cell proliferative disorder such as arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an infection, such as infections by viral agents classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and togavirus; infections by bacterial agents classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, and campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, and mycoplasma; infections by fungal agents classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malasezzia, histoplasma, and

other fungal agents causing various mycoses; and infections by parasites classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematodes such as trichinella, intestinal nematodes such as ascaris, lymphatic filarial nematodes, trematodes such as schistosoma, and cestrodes such as tapeworm. To assess abnormal expression of the present protein associated with any of these disorders, the level of the present polynucleotides or polypeptides can be detected in a biological sample or cell using any standard method, including Southern or northern analysis, dot blots, other membrane-based technologies, PCR technologies, dipstick, pin, ELISA assays, and in microarrays. Any of these methods may be used for the diagnosis of disorders characterized by an alteration of expression of SEQ ID NO:283 or 286, such as the disorders mentioned above, or in assays to monitor patients being treated with SEQ ID NO:283 or 286 or agonists, antagonists, or inhibitors of SEQ ID NO:283 or 286. Antibodies useful for diagnostic purposes may be prepared, e.g., in the same manner as that described in U.S. Patent No. 6,135,941. Diagnostic assays for SEQ ID NO:283 or 286 include methods which utilize the antibody and a label to detect SEQ ID NO: 283 or 286 in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

In another embodiment, the protein of SEQ ID NO:283 or 286 or a fragment or derivative thereof may be administered to a subject to diagnose, treat or prevent an immune disorder associated with decreased expression or activity of the protein of the invention. Such disorders can include, but are not limited to, acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, leukemias

such as multiple myeloma, and lymphomas such as Hodgkin's disease. In addition, such disorders associated with decreased protein expression or activity can be treated by administering to a patient polynucleotide sequences encoding the protein of the invention, e.g. inserted in an appropriate vector. In another example, a compound that increases either the activity of the protein of the invention or their expression can be administered to a patient to treat or prevent any of the diseases mentioned above.

In a further embodiment, an antagonist of the protein of the invention may be administered to a subject to treat or prevent an immune disorder associated with increased expression or activity of the protein of SEQ ID NO:283 or 286 including, but not limited to, auto-immune deseases or graft rejection. In one aspect, an antibody which specifically binds the protein of the invention may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express the proteins of the invention, such as the salivary gland tissue or the bone marrow tissue. In addition, sense, antisense nucleotides, GSE, ribozymes, specific protein inhibitors such as antibodies or small coumpounds can be administered to inhibit the expression of the proteins of the invention.

In another embodiment, an antagonist of the protein of SEQ ID NO:283 may be administered to a subject to treat or prevent a cell proliferative disorder. Such disorders may include, but are not limited to, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody which specifically binds the protein of the invention may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the protein of the invention. In another example, sense, antisense nucleotides, GSE, or ribozymes designed from nucleotides of the invention can be administered to inhibit the expression of the protein of the invention.

Protein of SEQ ID NO: 411 (internal designation 181-10-1-0-C9-CS)

The protein of SEQ ID NO: 411 encoded by the cDNA of SEQ ID No: 170 is highly expressed in fetal liver. The protein of the invention is homologous to peripheral benzodiazepine receptor/isoquinoline binding protein (PBR/IBP) of human, bovine and murine origin (Genbank accession numbers M36035, M64520 and L17306 respectively). The 170-amino-acid protein of SEQ ID NO: 411 is similar in size and hydropathicity to known peripheral PBR/IBP benzodiazepine receptors/isoquinoline binding proteins. Like the known peripheral benzodiazepine

receptors/isoquinoline binding proteins, the protein of the subject invention has about five potential transmembrane domains at positions 3-23, 45-65, 82-102, 105-125 and 130-150. Moreover, the protein of the invention displays a stretch of 11 amino acids (starting with V144 and ending with R154) that corresponds to a recently identified putative cholesterol recognition/interaction amino acid consensus pattern (-L/V-(X)(1-5)-Y-(X)(1-5)-R/K-) [See Li *et al*, *Endocrinology* 1998 Dec; 139 (12): 4991-7].

The peripheral benzodiazepine receptor (PBR) is a 18-kDa protein containing binding sites for benzodiazepine and is distinct from the GABA neurotransmitter receptor [Papadopoulos, V. (1993) *Endocr. Rev.* 14: 222-240]. Expression of PBR has been found in every tissue examined. However, it is most abundant in steroidogenic cells and is also found, primarily, on outer mitochondrial membranes [Anholt, R *et al.* (1986) *J. Biol. Chem.* 261:576-583]. PBR is thought to be associated with a multimeric complex composed of the 18-kDa isoquinoline binding protein and the 34-kDa pore-forming voltage dependent anion channel protein, preferentially located on the outer/inner mitochondrial membrane contact sites [McEnery, M.W. *et al. Proc. Natl. Acad. Sci.* USA. 89:3170-3174; Garnier, M. *et al.* (1994) *Mol. Pharmacol.* 45:201-211; Papadopoulos, V. *et al.* (1994) *Mol. Cel. Endocr.* 104:R5-R9]. Drug ligands of PBR, upon binding to the receptor, simulate steroid synthesis in steroidogenic cells in vitro [Papadopoulos, V *et al.* (1990) *J. Biol. Chem.* 265: 3772-3779; Barnea, E. R. *et al.* (1989) *Mol. Cell. Endocr.* 64: 155-159; Amsterdam, A. and Suh, B.S. (1991) *Endocrinology* 128: 503-510]. Likewise, *in vivo* studies showed that high affinity PBR ligands increase steroid plasma levels in hypophysectomized rats [Amri, H. *et al.* (1996) *Endocrinology* 137:5707-5718]. Further in vitro studies on isolated mitochondria provided evidence that PBR ligands, drug ligands, or the endogenous PBR ligand (the polypeptide diazepam-binding inhibitor (DBI) [Papadopoulos, V. *et al.* (1997) *Steroids* 62: 21-28]) stimulate pregnenolone formation by increasing the rate of cholesterol transfer from the outer to the inner mitochondrial membrane [for review, see Culty, M. *et al.* (1999) *Journal of Steroid Biochemistry and Molecular Biology* 69: 123-130].

Based on the amino acid sequence of the 18-kDa PBR, a three dimensional model was developed [Papadopoulos, V. (1996) In: The Leydig Cell. Payne, A. H. *et al.* (eds) Cache River Press, IL, pp 596-628]. This model was shown to accommodate a cholesterol molecule and function as a channel, supporting the role of PBR in cholesterol transport. The role of PBR in steroidogenesis was also demonstrated by observing that PBR negative cells generated by homologous recombination failed to produce steroids [Papadopoulos, V. *et al* (1997) *J. Biol. Chem.* 272: 32129-32135]. Further, cholesterol transport experiments in bacteria expressing the 18-kDa PBR protein provided definitive evidence for a function as a cholesterol channel/transporter [Papadopoulos, V. *et al.* (1997) supra].

In addition to its role in mediating cholesterol movement across membranes, PBR has been implicated in several other physiological functions, including cell growth and differentiation,

chemotaxis, mitochondrial physiology, porphyrin and heme biosynthesis, immune response, anion transport and GABAergic regulation of CNS. [for review, see Gavish, M. *et al.* (1999) *Pharmaceutical Reviews* 51: 629-650; Beurdeley-Thomas, A. *et al.* (2000) *Journal of Neuro-Oncology* 46: 45-56]. Also, a recent report also indicates that PBR agonists are potent anti-apoptotic compounds. These findings suggest that this effect may represent a major function for this receptor (Bono, F. *et al.* (1999) *Biochemical and Biophysical Research Communications* 265:457-461].

It appears that PBR is associated with stress and anxiety disorders. It has been suggested that PBRs play a role in the regulation of several stress systems such as the HPA axis, the sympathetic nervous system, the renin-angiotensin axis, and the neuroendocrine axis. In these systems, acute stress typically leads to increases in PBR density, whereas chronic stress typically leads to decreases in PBR density. Furthermore, in Generalized Anxiety Disorder (GAD), Panic Disorder (PD), Generalized Social Phobia (GSP), and Post-Traumatic Stress Disorders (PTSD), PBR density is typically decreased in platelets.

In the brain, where PBRs are associated with glial cells, PBRs are increased in specific brain areas in neurodegenerative disorders and also after neurotoxic and traumatic-ischemic brain damage [for review, see Gavish, M. *et al.* (1999) *supra*]. The literature also reports a decrease in peripheral-type benzodiazepine receptors in postmortems of chronic schizophrenics, suggesting that the decreased density of PBRs in the brain may be involved in the pathophysiology of schizophrenia. Increased levels of PBR in autopsied brain tissue from PSE patients (Portal-Systemic Encephalopathy patients) have been reported, thus supporting the theory that activation of PBR contributes to the pathogenesis characteristic of portal-systemic encephalopathy (PSE) in the central nervous system [Kurumaji, A. *et al.* (1997) *J. Neural Transm* 104:1361-1370; Butterworth R. F. (2000) *Neurochemistry International* 36: 411-416].

In addition to its involvement in the neurological disorders discussed supra, PBR has been implicated in the regulation of tumor cell proliferation [for review, see Gavish, M. *et al.* (1999) *supra*; Beurdeley-Thomas, A. *et al.* (2000) *supra*; Hardwick, M. (1999) *Cancer Research* 59:831-842; Venturini, I. *et al.* (1998) *Life Sci* 63:1269-80; Carmel I *et al.* (1999) *Biochem Pharmacol* 58: 273-8]. The invasiveness and metastatic ability of human breast tumor cells is proportional to the level of PBR expressed. Further, PBR has been proposed to be used as a tool/marker for detection, diagnosis, prognosis and treatment of cancer [WO 99/49316, hereby incorporated by reference in its entirety].

Many ligands have been described that bind to peripheral benzodiazepine receptor with various affinities. Some benzodiazepines, Ro 5-4864 [4-chlorodiazepam], diazepam and structurally related compounds, are potent and selective PBR ligands. Exogenous ligands also include 2-phenylquinoline carboxamides (PK11195 series), imidazo [1,2-a]pyridine-3-acetamides (Alpidem series) and pyridazine derivatives. Some endogenous compounds, including porphyrins and

diazepam binding inhibitor (DBI), bind to PBR with nanomolar and micromolar affinity [for review, see Gavish, M. *et al.* (1999) *supra*; Beurdeley- Thomas, A. *et al.* (2000) *supra*].

The protein of SEQ ID NO: 411 is a novel peripheral-type benzodiazepine receptor. As such, it is serves a channel function that mediates cholesterol movement across membranes, play a role in steroidogenesis, cell growth and differentiation, chemotaxis, mitochondrial physiology, protection against apoptosis, porphyrin and heme biosynthesis, immune response, anion transport and GABAergic regulation of CNS.

In one embodiment, a preferred polypeptide of the invention comprises the amino acids of SEQ ID NO: 411 from position 144 to 154. In another embodiment, the subject invention provides a polypeptide comprising the sequence of SEQ ID NO: 411. Other preferred polypeptides of the invention include biologically active fragments of SEQ ID NO: 411. Biologically active fragments of the protein of SEQ ID NO: 411 have any of the biological activities described herein which are associated with the PBR. In another embodiment, the polypeptide of the invention is encoded by clone 181-10-1-0-C9-CS.

One aspect of the subject invention provides compositions and methods using the protein of the invention, or biologically active fragments thereof, for the development, identification, and/or selection of agents capable of modulating the expression or activity of the protein of the invention.

Agents which modulate the activity of the PBR/IBP of the subject invention include, but are not limited to, antisense oligonucleotides, ribozymes, drugs, and antibodies. These agents may be made and used according to methods well known in the art. Also, the protein of the invention, or biologically active fragments thereof, may be used in screening assays for therapeutic compounds. A variety of drug screening techniques may be employed. In this aspect of the invention, the protein or biologically active fragment thereof, may be free in solution, affixed to a solid support, recombinantly expressed on, or chemically attached to, a cell surface, or located intracellularly. The formation of binding complexes, between the protein of the invention, or biologically active fragments thereof, and the compound being tested, may then be measured.

In one embodiment, the subject method utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the PBR/IBP polypeptide or biologically active fragments thereof. The transformed cells may be viable or fixed. Drugs or compounds which are candidates for the modulation of the PBR/IBP, or biologically active fragments thereof, are screened against such transformed cells in binding assays well known to those skilled in the art. Alternatively, assays such as those taught in Geysen H. N., WO Application 84/03564, published on Sep. 13, 1984, and incorporated herein by reference in its entirety, may be used to screen for peptide compounds which demonstrate binding affinity for, or the ability to modulate, the PBR/IBP, or biologically active fragments thereof. In another embodiment, competitive drug screening assays using neutralizing antibodies specifically compete with a test

compound for binding to the PBR/IBP protein of the invention, or biologically active fragments thereof.

Another embodiment of the subject invention provides compositions and methods of selectively modulating the expression or activity of the protein of the invention. Modulation of the PBR/IBP would allow for the successful treatment and/or management of diseases or biochemical abnormalities associated with the PBR or PBR/IBP. Antagonists, able to reduce or inhibit the expression or the activity of the protein of the invention, would be useful in the treatment of diseases associated with elevated levels of the PBR/IBP, increased cell proliferation, or increased cholesterol transport. Thus, the subject invention provides methods for treating a variety of diseases or disorders, including, but not limited to, cancers, especially liver cancer, and portal-systemic encephalopathy.

Alternatively, the subject invention provides methods of treating diseases or disorders associated with decreased levels of the protein of the PBR/IBP. Thus, the subject invention provides methods of treating diseases including, and not limited to, schizophrenia, chronic stress, GAD, PD, GSP and PTSD. Other diseases which may be treated by agonists of the PBR/IBP of the subject invention include those diseases associated with decreases in cell proliferation, e.g. developmental retardation.

Furthermore, because the PBR/IBP of the subject invention is also able to transport cholesterol into cells, the subject invention may also be used to increase cholesterol transport into cells. Diseases associated with cholesterol transport deficiencies include lipoidal adrenal hyperplasia, and diseases where there is a requirement for increased production of compounds requiring cholesterol such as myelin and myelination, such as Alzheimer's disease, spinal chord injury, and brain development neuropathy [Snipes, G. and Suter, U. (1997) Cholesterol and Myelin. In: Subcellular Biochemistry, Robert Bittman (ed.), vol. 28, pp. 173-204, Plenum Press, New York]. The methods of treating disorders associated with decreased levels of PBR/IBP may be practiced by introducing agonists which stimulate the expression or the activity of the protein of the invention.

In one embodiment, methods of increasing the levels of PBR/IBP in tissues or cell types may be practiced by utilizing nucleic acids encoding the protein of the subject invention, or biologically active fragments thereof, to introduce biologically active polypeptide into targeted cell types. Vectors useful in such methods are known to those skilled in the art as are methods of introducing such nucleic acids into target tissues.

Agents which stimulate or inhibit the activity of the protein of the invention include but are not limited to agonist and antagonist drugs respectively. These drugs can be obtained using any of a variety of drug screening techniques as discussed above.

Antagonists of the PBR/IBP encoded by SEQ ID NO: 170 include agents which decrease the levels of expressed mRNA encoding the protein of SEQ ID NO: 411. These include, but are not limited to, RNAi, one or more ribozymes capable of digesting the protein of the invention mRNA,

or antisense oligonucleotides capable of hybridizing to mRNA encoding the PBR/IBP of SEQ ID NO: 411 Antisense oligonucleotides can be administrated as DNA, as DNA entrapped in proteoliposomes containing viral envelope receptor proteins [Kanoda, Y. *et al.* (1989) *Science* 243: 375] or as part of a vector which can be expressed in the target cell and provide antisense DNA or RNA. Vectors which are expressed in particular cell types are known in the art. Alternatively, the DNA can be injected along with a carrier. A carrier can be a protein such as a cytokine, for example interleukin 2, or polylysine-glycoprotein carriers. Carrier proteins, vectors, and methods of making and using polylysine carrier systems are known in the art. Alternatively, nucleic acid encoding antisense molecules may be coated onto gold beads and introduced into the skin with, for example, a gene gun [Ulmer, J.B. *et al.* (1993) *Science* 259:1745].

Antibodies, or other polypeptides, capable of reducing or inhibiting the activity of PBR/IBP may be provided as in isolated and substantially purified form. Alternatively, antibodies or other polypeptides capable of inhibiting or reducing the activity of the PBR/IBP protein, may be recombinantly expressed in the target cell to provide a modulating effect. In addition, compounds which inhibit or reduce the activity of the PBR/IBP protein of the subject invention may be incorporated into biodegradable polymers being implanted in the vicinity of where drug delivery is desired. For example, biodegradable polymers may be implanted at the site of a tumor or, alternatively, biodegradable polymers containing antagonists/agonists may be implanted to slowly release the compounds systemically. Biodegradable polymers, and their use, are known to those of skill in the art (see, for example, Brem *et al.* (1991) J. *Neurosurg.* 74:441-446.

In another embodiment, the invention provides methods and compositions for detecting the level of expression of the mRNA of the protein of the invention. Quantification of mRNA levels of the PBR/IBP protein of the invention may be useful for the diagnosis or prognosis of diseases associated with an altered expression of the protein of the invention. Assays for the detection and quantification of the mRNA of the protein of the invention are well known in the art (see, for example, Maniatis, Fitsch and Sambrook, Molecular Cloning; A Laboratory Manual (1982), or Current Protocols in Molecular Biology, Ausubel, F.M. *et al.* (Eds), Wiley & Sons, Inc.).

Polynucleotides probes or primers for the detection of the mRNA of the protein of SEQ ID NO: 411 can be designed from the cDNA of SEQ ID NO: 170. Methods for designing probes and primers are known in the art. In another embodiment, the subject invention provides diagnostic kits for the detection of the mRNA of the protein of the invention in cells. The kit comprises a package having one or more containers of oligonucleotide primers for detection of the protein of the invention in PCR assays or one or more containers of polynucleotide probes for the detection of the mRNA of the protein of the invention by in situ hybridization or Northern analysis. Kits may, optionally, include containers of various reagents used in various hybridization assays. The kit may also, optionally, contain one or more of the following items: polymerization enzymes, buffers, instructions, controls, or detection labels. Kits may also, optionally, include containers of reagents

mixed together in suitable proportions for performing the hybridization assay methods in accordance with the invention. Reagent containers preferably contain reagents in unit quantities that obviate measuring steps when performing the subject methods.

In another embodiment, the invention relates to methods and compositions for detecting and quantifying the level of the protein of the invention present in a particular biological sample. These methods are useful for the diagnosis or prognosis of diseases associated with an altered levels of the protein of the invention. Diagnostic assays to detect the protein of the invention may comprise a biopsy, in situ assay of cells from organ or tissue sections, or an aspirate of cells from a tumor or normal tissue. In addition, assays may be conducted upon cellular extracts from organs, tissues, cells, urine, or serum or blood or any other body fluid or extract.

Assays for the quantification of the PBR/IBP of SEQ ID NO: 411 may be performed according to methods well known in the art. Typically, these assays comprise contacting the sample with a ligand of the protein of the invention or an antibody (polyclonal or monoclonal) which recognizes the protein of the invention or a fragment thereof, and detecting the complex formed between the protein of the invention present in the sample and the ligand or antibody. Fragments of the ligands and antibodies may also be used in the binding assays, provided these fragments are capable of specifically interacting with the BRP/IRP of the subject invention. Further, the ligands and antibodies which bind to the BRP/IRP of the invention may be labeled according to methods known in the art. Labels which are useful in the subject invention include, but are not limited to, enzymes labels, radioisotopic labels, paramagnetic labels, and chemiluminescent labels. Typical techniques are described by Kennedy, J. H., *et al.* (1976) *Clin. Chim. Acta* 70:1-31; and Schurs, A. H. *et al.* (1977) *Clin. Chim. Acta* 81: 1-40.

The subject invention also provides methods and compositions for the identification of metastatic tumor masses. In this aspect of the invention, the polypeptides and antibodies which bind the polypeptides of the invention may be used as a marker for the identification of the metastatic tumor mass. Metastatic tumors which originated from the liver may overexpress the PBR/IBP of SEQ ID NO: 411, whereas newly forming tumors, or those originating from other tissues are not expected to bear the PBR/IBP of SEQ ID NO: 411.

Protein of SEQ ID NO: 397 (internal designation 160-28-4-0-C4-CS).

The protein of SEQ ID NO: 397, encoded by the cDNA of SEQ ID NO: 156 (clone 160-28-4-0-C4-CS), exhibits homology to the ADP-ribosylation factors (ARF) family of proteins. The ARF family includes ADP-ribosylation factors (ARFs) and ARF-like proteins (ARLs); the ARF family of proteins is one family of the Ras superfamily. Proteins belonging to the Ras superfamily have molecular weights of 18-30 kDa and function in a variety of cellular processes including, but not limited to, signaling, growth, immunity, and protein transport.

ARFs are monomeric GTP-binding proteins, related structurally to both G protein alpha-subunits and Ras proteins. ARF family members share more than 60% sequence identity, appear to be ubiquitous in eukaryotes, and are evolutionarily highly conserved throughout. Immunologically, they have been localized to the Golgi apparatus of several types of cells (Stearns et al. Proc. Natl. Acad. Sci. (USA) 87:1238-1242 (1990)). ARF proteins enhance the ADP-ribosyltransferase activity of cholera toxin as an allosteric activator (Noda et al. Biochim. Biophys. Acta 1034: 195-199 (1990)). ARFs have also been shown to act as regulatory molecules, or "switches", for linking two processes (*e.g.,* the process of vesicle fission from a donor compartment and fusion with an acceptor compartment (Rothman, J. E. and Wieland, F. T. Science 272: 227-234 (1996)). ARF family members fall into three classes, classes I-III, according to their size and sequence homology. Class I comprises ARF1, ARF2, and ARF3; Class II comprises ARF4 and ARF5; and Class III comprises ARF6.

The classes occupy different subcellular locations and have been implicated in different transport pathways. Class I ARFs localize to the Golgi where they are involved in the regulation of ER-Golgi and intra-Golgi transport. Class I ARFs are also involved in the recruitment of cytosolic coat proteins to Golgi membranes during the formation of transport vesicles. Class III (*e.g.*, ARF6) localizes to a tubulovesicular compartment, secretory granules, and the plasma membrane, where it is involved in regulated secretion and recycling. Class II ARFs appear to be cytosolic, but their role has not been elucidated. (Radhakrishna, H. and Donaldson, J. G. J. Cell Biol. 139: 49-61(1997)).

ARF function, in general, is regulated by a GDP-GTP cycle. For example, ARF1 is cytosolic in the GDP bound state, but is associated with membranes when in the GTP bound state. A guanine nucleotide exchange factor (GEF) in the donor compartment recruits ARF1 to the membrane. At the membrane, GTP-ARF1 recruits coat proteins, which assemble together into spherical coats, budding off vesicles in the process. After budding, hydrolysis of bound GTP causes ARF1 to dissociate from the membrane. ARF1 dissociation causes the coat to become unstable and dissociate as well. (Rothman, supra.)

Members of the ARF multigene family, when expressed as recombinant proteins in *E. coli*, display different phospholipid and detergent requirements (Price, et al. J. Biol. Chem. 267: 17766-17772 (1992)). Some lipids and/or detergents, e.g., SDS, cardiolipin, dimyristoylphosphatidylcholine (DMPC)/cholate, enhance ARF activities (Bobak, et al. Biochemistry 29:855-861 (1990); Noda, et al. Biochim. Biophys. Acta 1034: 195-199 (1990); Tsai, et al. J. Biol. Chem. 263:1768-1772 (1988)). ARFs also activate phospholipase D (PLD), a membrane-bound enzyme implicated as an effector of several growth factors (Boman, A. L. and Kahn, R. A. Trends Biochem. Sci. 20: 147-150 (1995). PLD1 has been shown to be activated by a variety of G-protein regulators, for example, PKC (protein kinase C) and ADP-ribosylation factor (ARF). PKC and ARFs may regulate G-proteins either individually or together in a synergistic manner. Recently the role of ARFs in microtubules formation has also been demonstrated. ADP-

ribosylation of tubulin almost completely blocked self-assembly of this protein in brain (Terashima M. et a; J.Nutr Sci Vitaminol 45: 393-400 (1999)).

In general, differences in the various ARF sequences are concentrated in the amino-terminal regions and the carboxyl portions of the proteins. Only three of 17 amino acids in the amino termini have shown to be identical among ARFs, and four amino acids in this region of ARFs 1-5 are missing in ARF 6 (Tsuchiya, et al. J. Biol. Chem. 266: 2772-2777 (1991)). It was reported (Kahn, et al. J. Biol. Chem. 267:13039-13046 (1992)) that the amino-terminal regions of ARF proteins form an alpha-helix and that this domain is required for membrane targeting, interaction with lipid, and ARF activity.

Schliefer et al., (J. Biol. Chem. 257: 20-23 (1991)) have described a protein distinctly larger than ARF that possessed ARF-like activity. ARF-like proteins, or ARLs, have been found in different species. Some of ARLs appear to lack ADP-ribosyltransferase-enhancing activity; ARLs may differ in GTP-binding requirements and GTPase activity as compared to various ARF isoforms. For example, ARP, a mammalian ARL, is 33-39% identical to members of the ARF family; ARP, however, differs from other ARF family proteins by virtue of its ability to hydrolyze bound GTP in the absence of other proteins. ARP protein, unlike ARFs, is typically associated with plasma membrane instead of the cytosol (Schurmann, A. J. Biol. Chem. 270, 30657-30663 (1995)).

ARF family members have been implicated in several disease processes, such as Lowe's syndrome, an X-linked disorder characterized by congenital cataracts, renal tubular dysfunction and neurological deficits. These disorders may be due to an inability to recruit ARF to the Golgi membrane (Suchy, S. F. et al. Hum. Mol. Genet. 4: 2245-2250 (1995), Londono I. et al. Kidney Int. 55: 1407-1416 (1999)). It has also been suggested that regulation of ARF is also involved in cystic fibrosis, Dent's disease, diabetes, and autosomal dominant polycystic kidney disease (Marshansky, V., et al. Electrophoresis 18: 2661-2676 (1997)).

The new human ARF-related protein of SEQ ID NO:397, encoded by clone 160-28-4-0-C4-CS in one embodiment, and the related polynucleotides, provide new compositions which are useful in the diagnosis, treatment, and prevention of secretory, exocytosis, endocytosis and another "sorting disorders."

The subject invention provides a polypeptide comprising the amino acid sequence of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, or biologically active fragments thereof. The intact protein of interest is 173 amino acids in length, has an ARF family amino acid motif (Pfam), and has ATP/GTP-binding site motif A P-loop (PS00017). The protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS also has chemical and structural similarity with human ARL1 (P40616), ARD-1 (R66033) and ARF6 (GI 178989) (31%, 31% and 27% identity, respectively). The amino acid length of SEQ ID NO: 397 is similar to those of the aforementioned ARFs Biologically active fragments of SEQ ID NO: 397 have one or more of the biological activities typically associated the full length protein. In one embodiment, the protein is encoded by clone 160-28-4-0-C4-CS

The invention also provides variants of the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS. The variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS. Variants according to the subject invention have at least one functional and/or structural characteristic of ARFs. The invention also provides biologically active fragments of the variant proteins.

The invention includes those polynucleotides encoding the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, variants of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, and biologically active fragments of both the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS and variants thereof. As is apparent to those skilled in the art, a variety of different DNA sequences can encode the amino acid sequence of the proteins, variants, and biologically active fragments of said proteins and variants. It is well within the skill of a person trained in the art to create these alternative DNA sequences encoding proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are also within the scope of the subject invention. As used herein, reference to "essentially the same" sequence refers to sequences that have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity.

The subject invention provides method of treating cytoskeletal, secretory, and inflammatory disorders/conditions comprising the administration of therapeutically effective amounts of a composition comprising the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS. These methods can also be practiced using variants of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, or biologically active fragments of either SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, or variants of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS. Disorders/conditions which can be treated by the subject invention include, but are not limited to, prostate cancer, brain and another tumors, Lowe's syndrome, glomerulonephritis, chronic glomerulonephritis, tubulointerstitial nephritis, inherited X-linked nephrogenic diabetes insipidus, autosomal dominant polycystic kidney disease (ADPKD), herpes gestationis, dermatitis herpetiformis, lupus erythematosus, Crohn's disease, irritable bowel syndrome and Addison's disease; secretory/endocytotic disorders such as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, hyper- and hypoglycemia, Grave's disease, goiter, and Cushing's disease; conditions associated with abnormal vesicle trafficking, including acquired immunodeficiency syndrome (AIDS); allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia; multiple sclerosis; myasthenia gravis; rheumatoid and osteoarthritis; Chediak-Higashi and Sjogren's syndromes; toxic shock syndrome; traumatic tissue damage; viral, bacterial, fungal, helminthic, and protozoal infections.

In another embodiment, a vector capable of expressing the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, or biologically active fragments thereof, can be administered to a subject to treat or prevent disorders including, but not limited to, those described above. Alternatively, the vector can encode a variant, or biologically active fragment of the variant protein. Multiple vectors

encoding any combination of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, variants, and/or biologically active fragments of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS and/or variants can be administered to a subject.

In a further embodiment, a pharmaceutical composition comprising a substantially purified protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS (and/or biologically active fragments thereof), in conjunction with a suitable pharmaceutical carrier, can be administered to a subject to treat or prevent the above mentioned disorders. Alternatively, a pharmaceutical composition comprising a substantially purified variant protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS (and/or biologically active fragments thereof), in conjunction with a suitable pharmaceutical carrier, can be administered in the aforementioned therapeutic regimens. As would be apparent to the skilled artisan, any therapeutically effective combination of the protein encoded by SEQ ID NO: 397 or clone 160-28-4-0-C4-CS (and/or biologically active fragments thereof) and variants of SEQ ID NO:397 or clone 160-28-4-0-C4-CS (and/or biologically active fragments thereof), in conjunction with a suitable pharmaceutical carrier can be used in the aforementioned therapeutic regimens.

ARFs are known to be involved in regulated transport of vesicles. Therefore, in another embodiment, the protein of SEQ ID No: 397 or clone 160-28-4-0-C4-CS, variants, and/or biologically active fragments of said proteins and/or variants can be used as a component of drug delivery vehicles such as colloids or liposomes. The protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, variants, and/or biologically active fragments of said proteins and/or variants can be incorporated into the lipid membranes of liposomes and can serve as specific targeting agents. The methods of design of such drug delivery systems is known by those skilled in the art and can be practiced according to conventional pharmaceutical principles (Smith H.J. Introduction to the principles of drug design and action, 3rd ed. (1998); Chien Y.W. Novel Drug Delivery systems, 2nd ed. (1992); Storm G. et al J.Liposome Res. 4: 641-666 (1994); and Crommelin D.J.A. et al. Adv. Drug Delivery Rev. 17 : 49-60 (1995)).

In another embodiment of the invention, the polynucleotides encoding the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS can be used for therapeutic purposes. Polynucleotides encoding fragments of the protein of SEQ ID NO:397 or clone 160-28-4-0-C4-CS can also be used in therapeutic regiments. In one aspect, the complement of the polynucleotide encoding the protein of SEQ ID NO.: 397 or clone 160-28-4-0-C4-CS can be used in situations in which it would be desirable to block the transcription of the mRNA. Modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding the protein of interest. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding the

protein of interest. Methods of treatment utilizing antisense technology are also well known to those skilled in the art.

Another embodiment of the invention provides methods of assessing PLD modulation by using ARF properties of the protein of interest.

In another embodiment, antibodies which specifically bind the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS can be used for the diagnosis of disorders characterized by expression of the protein, or in assays to monitor patients being treated with the protein of interest. Methods of making both polyclonal and monoclonal antibodies are well-known in the art. Diagnostic assays which can be used in this aspect of the invention include, and are not limited to, ELISAs, RIAs, and FACS, and are well known in the art. These assays also provide a basis for diagnosing or identifying altered or abnormal levels of SEQ ID NO:397 or the polypeptides encoded by the human cDNA of clone 160-28-4-0-C4-CS expression as compared to normal individuals. These screening methods are, likewise, well known to the skilled artisan.

In another embodiment of the invention, the protein of interest, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening can be free in solution, affixed to a solid support, recombinantly expressed on, or chemically attached to, a cell surface, or located intracellularly. The formation of binding complexes between the protein of interest and the agent being tested can be measured by methods well known to those skilled in the art. Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, *e.g.*, Geysen, et al. (1984) PCT application WO84/03564.)

In another embodiment of the invention, the polynucleotides encoding the protein of interest can be used for diagnostic purposes. The polynucleotides can be used to detect and quantify gene expression in biopsied tissues in which expression of the protein of interest can be correlated with a disease or condition. Such diagnostic assays are well known in the art and can be used to monitor regulation of the protein of interest levels during therapeutic intervention and/or to determine absence, presence, and excess expression of the protein of interest. Examples of such conditions and disorders have been provided supra. The polynucleotide sequences encoding the protein of interest can be used, for example, in Southern or Northern analyses, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered expression of the protein of SEQ ID NO:397 or clone 160-28-4-0-C4-CS. Such qualitative or quantitative methods are well known in the art.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein can be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to

identify genetic variants, mutations, and polymorphisms. This information can be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents. Microarrays can be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T. M. et al. (1995) U.S. Pat. No. 5,474,796; Heller, R. A. et al. (1997) Proc. Natl. Acad. Sci. 94: 2150-2155; and Heller, M. J. et al. (1997) U.S. Pat. No. 5,605,662.)

Another embodiment of the subject invention provides nucleic acid sequences encoding the protein of interest which can be extended utilizing a partial nucleotide sequence and various PCR-based methods. This aspect of the invention provides methods for the detection of upstream sequences, such as promoters and regulatory elements. Methods of practicing this aspect of the invention are also well known in the art.

In other embodiments of the disclosed therapeutic regimens, any of the proteins, variants, biologically active fragments, antibodies, complementary sequences, or vectors of the invention can be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy can be made by one of ordinary skill in the art. The combination of therapeutic agents can act synergistically to effect the treatment or prevention of the various disorders described above. In particular, purified protein can be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind the protein of interest. Neutralizing antibodies especially preferred for therapeutic use.

Protein of SEQ ID NO: 287 (internal designation 174-5-3-0-H7-CS)

The protein of SEQ ID NO: 287, encoded by human cDNA of SEQ ID NO: 46 (clone 174-5-3-0-H7-CS), is highly homologous (more than 99% identity in amino acids) to the human protein encoded by the CLN8 gene listed in Genbank under accession number AF123757. The two proteins differ by two conservative amino-acid substations (alanine for valine at position 155 and serine for asparagine at position 225). In addition, the protein encoded by 174-5-3-0-H7-CS contains seven transmembrane domains. These domains are located at amino acids 25-45, 71-91, 100-120, 133-153, 160-180, 205-225, and 228-248 as predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, 10:685-686 (1994)). The protein encoded by SEQ ID NO: 287 also exhibits a signal peptide at positions 1-50 and a retention signal KKRP from positions 283 to 286.

CLN8 was identified recently by positional cloning (Ranta et al., Nat Genet. 1999 Oct.;23(2):233-6). CLN8 encodes a 286 amino-acid putative transmembrane protein with no homology to previously known proteins. A naturally-occurring missense mutation in codon 24 (R24G at the border of the first putative transmembrane domain) is the molecular basis for EPMR ("progressive epilepsy with mental retardation", MIM 600143). EPMR, also called Northern Epilepsy, is an autosomal recessive disorder characterized by normal early development, onset of

generalized tonic-clonic seizures between the ages of 5 and 10 years, and subsequent progressive mental retardation. Neuropathological findings have shown that EPMR is a new member of the neuronal ceroid lipofuscinosis (NCL) group of neurodegenerative disorders. The NCLs are a genetically heterogeneous group of progressive neurodegenerative disorders characterized by the accumulation of autofluorescent lipopigment in various tissues. CLN8 is the eighth gene to be linked to the NCL group of neuro-degenerative disorders.

Subsequently, the homologous mouse gene (Cln8) was sequenced (82% nucleotide identity with the human gene) and localized to the region of the mouse genome linked to motor neuron degeneration, mouse mnd. Mnd is a naturally-occurring mouse mutant with intracellular autofluorescent inclusions similar to those seen in EPMR. A mutation in mnd mouse DNA was identified, indicating that mnd is a murine ortholog for CLN8 (Ranta et al., Nat Genet. 1999 Oct;23(2):233-6), and that mice containing mutations in Cln8 represent a murine model for NCL disorders.

Recent experimental evidence has confirmed the transmembrane nature of the CLN8 protein (Lonka L et al., Hum Mol Genet. 2000 Jul 1;9(11):1691-7). CLN8 resides in the endoplasmic reticulum (ER) and recycles between the ER and the ER-Golgi intermediate compartment (ERGIC) via a KKXX ER-retrieval motif at its C-terminus (KKRP, amino-acids 283-286). This motif is recognized and bound by COPI, a vesicle-coating protein found in retrograde vesicles delivering cargo from the *cis* Golgi to the ER. The 30kD CLN8 protein is not processed during its maturation (in particular it is not N-glycosylated). The EPMR-associated R24G mutation does not alter cellular localization in humans.

The subject invention provides a polypeptide encoded by SEQ ID NO: 287 and biologically active fragments of said polypeptide. Compositions comprising polypeptides and pharmaceutically acceptable carriers are likewise provided. Preferred polypeptides, and biologically active fragments thereof, have any of the biological activities or domains/motifs described herein and/or contain the amino acids of positions 155 and 225, 283 to 286. In one embodiment, the protein/polypeptide of SEQ ID NO: 287 is encoded by clone 174-5-3-0-H7-CS.

The ER/ERGIC cellular localization of protein of this invention can be used to target compounds to the ER/ERGIC. This targeting can be observed using any of the techniques known to those skilled in the art including those described in (Lonka L et al., Hum Mol Genet. 2000 Jul 1;9(11):1691-7). In this aspect of the invention, the protein of SEQ ID NO: 287, or biologically active fragments thereof can be used to target liposomes, vesicles, or colloids to the ER/ERGIC compartment where active agents can be delivered. Methods of making and using targeted liposomes are well known in the art.

In another embodiment, liposomes comprising the protein of SEQ ID NO: 287 can contain a second targeting agent for the specific selection of a target cell. The second targeting agent can be selected for its ability to specifically target a cell or tissue. Thus, the second targeting agent can be

specific for tumor markers, such as HER2. Alternatively, markers associated with specific cell types can be used (e.g., CD34, CD4, CD8, etc.). In a preferred embodiment, the second targeting agent is an antibody. Active agents include, but are not limited to, chemotherapeutic agents protein cross-linking agents, inhibitors of protein synthesis, anti-bacterial agents (e.g., antibiotics), antiviral agents, and/or anti-parasitic agents. The ability to bind the COPI coatomer can be assayed as described in (Cosson P, Letourneur F, Science. 1994 Mar 18;263(5153):1629-31).

In another embodiment, the present invention provides methods of, and compositions for, identifying specific cellular compartments, such as the ER, ERGIC, and retrograde transport vesicles. This embodiment provides antibodies which specifically bind the protein of SEQ ID NO: 287, or biologically active fragments thereof, which are labeled with detectable markers, such as gold particles, enzymes, radioisotopes, or paramagnetic labels. ER, ERGIC, and retrograde transport vesicles can be identified in samples according to well-known immuno-diagnostic protocols. The antibodies, either monoclonal or polyclonal, can be made according to well-known methods. In a preferred embodiment, the antibodies bind to ER retention signal.

In another embodiment, the protein of the invention or part thereof can be used as a reagent for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, asthma, pulmonary edema, atherosclerosis, restenosis, stroke potential, thrombosis and hypertension. Similarly, the protein of the invention, or biologically active fragments thereof, and antibodies thereto can provide immunological probes for differential identification of the tissue(s) or cell type(s). In a number of disorders listed above, particularly of the pulmonary and cardiovascular systems, expression of this protein at significantly higher or lower levels can be routinely detected in certain tissues or cell types (e. g., vascular tissues, cancerous and wounded tissues) or bodily fluids (e. g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Indeed, the 80 first amino-acids of the protein of the invention are identical to two polypeptides claimed in Patent WO 99/35158, hereby incorporated by reference in its entirety (SEQ ID NO:98 and SEQ ID NO:162 corresponding to Geneseq accession numbers Y38413/Y38428 and Y38492) are over-expressed in pulmonary and endothelial tissues.

The tissue distribution in pulmonary and endothelial tissues indicates that the protein product described in WO 99/35158 is useful for the treatment and diagnosis of cardiovascular and respiratory or pulmonary disorders such as asthma, pulmonary edema, pneumonia, atherosclerosis, restenosis, stroke, angina, thrombosis hypertension, inflammation, and wound healing. Those conditions can be diagnosed by determining the amount of the protein of the invention in a sample. Thus, antibodies raised against the protein of SEQ ID NO: 287, or an immunogenic fragment of the

protein can be used in diagnostic, prognostic, or screening assays such as those taught in WO 99/35158.

Protein of SEQ ID No. 270 (internal designation 116-119-3-0-H5-CS)

The protein of SEQ ID NO: 270 encoded by the extended cDNA SEQ ID NO: 29 is homologous to the human mitochondrial ATP synthase f subunit or ATPK (E.C. 3.6.1.34) (Swissprot accession number P56134) and is overexpressed in fetal kidney.

The protein of SEQ ID NO: 270, composed of 88 amino acid residues, contains 1 transmembrane segment (position 1 to 55) predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, 10 :685-686 (1994). BLAST results show that 100% homology is found between amino acids 5 to 88 of the protein of the invention and amino acids 10 to 93 of human ATP synthase f chain (93 amino acids total), exon 1 of the cDNA SEQ ID NO: 29 making the difference between the 2 proteins (the last 3 exons show 100% homology). Thus, the protein of the invention represents a new isoform of human mitochondrial ATP synthase f subunit. It is interesting to note that the same splice variant is found in bovin, pig and mouse species.

The mitochondrial electron transport (or respiratory) chain is a series of enzyme complexes in the mitochondrial membrane that is responsible for the transport of electrons from NADH to oxygen and the coupling of this oxidation to the synthesis of ATP (oxidative phosphorylation). ATP then provides the

primary source of energy for driving a cell's many energy-requiring reactions. ATP synthase (F0 F1 ATPase) is the enzyme complex at the terminus of this chain and serves as a reversible coupling device that interconverts the energies of an electrochemical proton gradient across the mitochondrial membrane into either the synthesis or hydrolysis of ATP. This gradient is produced by other enzymes of the respiratory chain in the course of electron transport from NADH to oxygen. When the cell's energy demands are high, electron transport from NADH to oxygen generates an electrochemical gradient across the mitochondrial membrane. Proton translocation from the outer to the inner side of the membrane drives the synthesis of ATP. Under conditions of low energy requirements and when there is an excess of ATP present, this electrochemical gradient is reversed and ATP synthase hydrolyzes ATP. The energy of hydrolysis is used to pump protons out of the mitochondrial matrix. ATP synthase is, therefore, a dual complex, the F0 portion of which is a transmembrane proton carrier or pump, and the F1 portion of which is catalytic and synthesizes or hydrolyzes ATP. Mammalian ATP synthase complex consists of sixteen different polypeptides (Walker, J. E. and Collinson, T. R. (1994) FEBS Lett.346: 39-43). Six of these polypeptides (subunits alpha, beta, gamma, delta, epsilon, and an ATPase inhibitor protein IF 1) comprise the globular catalytic F 1 ATPase portion of the complex, which lies outside of the mitochondrial membrane. The remaining ten polypeptides (subunits a, b, c, d, e, f, g, F6, OSCP, and A6L) comprise the proton-translocating, membrane spanning F0 portion of the complex. Like other members of the respiratory chain, all but two of the polypeptide subunits of ATP synthase

are nuclear gene products that are imported into the mitochondria. Enzyme complexes similar to mammalian ATP synthase are found in all cell types and in chloroplast and bacterial membranes. This universality indicates the central importance of this enzyme to ATP metabolism. Transcriptional regulation of these nuclear encoded genes appears to be the predominant means for controlling the biogenesis of ATP synthase. Multiple mitochondrial pathologies exist because of the essential role of mitochondrial oxidative phosphorylation in cellular energy production, in the generation of reactive oxygen species and in the initation of apoptosis (Wallace, Science, 283:1482-1488, 1999). It is now clear that mitochondrial diseases encompass an assemblage of clinical problems commonly involving tissues that have high energy requirements such as heart, muscle and the renal and endocrine systems. Over the past 11 years, a considerable body of evidence has accumulated implicating defects in the mitochondrial energy-generating pathway, oxidative phosphorylation, in a wide variety of degenerative diseases including myopathy and cardiomyopathy. Most classes of pathogenic mitochondrial DNA mutations affect the heart, in association with a variety of other clinical manifestations that can include skeletal muscle, the central nervous system (including eye), the endocrine system, and the renal system. Nuclear mutations causing mitochondrial disorders have been described. They are often found in highly conserved subunits. Mitochondrial disorders with nuclear mutations include : myopathies (PEO, MNGIE, congenital muscular dystrophy, carnitine disorders), encephalopathies (Leigh, Infantile, Wilson's disease, Deafness-Dystonia syndrome), other systemic disorders and cardiomyopathies.

The discovery of a new ATP synthase subunit, and polynucleotides encoding it satisfy a need in the art by providing new compositions which are useful for the diagnosis, prevention, and treatment of cancer, myopathies, immune disorders, and neurological disorders.

It is believed that the protein of SEQ ID NO: 270 or part thereof plays a role in cellular respiration, preferably as a mitochondrial ATP synthase subunit. Preferred polypeptides of the invention are fragments of SEQ ID NO: 270 having any of the biological activity described herein.

An object of the present invention are compositions and methods of targeting heterologous compounds, either polypeptides or polynucleotides to mitochondria by recombinantly or chemically fusing a fragment of the protein of the invention to an heterologous polypeptide or polynucleotide. Preferred fragments are signal peptide, amphiphilic alpha helices and/or any other fragments of the protein of the invention, or part thereof, that may contain targeting signals for mitochondria including but not limited to matrix targeting signals as defined in Herrman and Neupert, Curr. Opinion Microbiol. 3:210-4 (2000); Bhagwat et al. J. Biol. Chem. 274:24014-22 (1999), Murphy Trends Biotechnol. 15:326-30 (1997); Glaser et al. Plant Mol Biol 38:311-38 (1998); Ciminale et al. Oncogene 18:4505-14 (1999). Such heterologous compounds may be used to modulate mitochondria's activities. For example, they may be used to induce and/or prevent mitochondrial-induced apoptosis or necrosis. In addition, heterologous polynucleotides may be used for mitochondrial gene therapy to replace a defective mitochondrial gene and/or to inhibit the deleterious expression of a mitochondrial gene.

The invention further relates to methods and compositions using the protein of the invention or part thereof to diagnose, prevent and/or treat several disorders in which mitochondrial respiratory electron transport chain is impaired, including but not limited to mitochondriocytopathies, necrosis, aging, myopathies, cancer and neurodegenerative diseases such as Alzheimer's disease, Huntington's disease, Parkinson's disease, epilepsy, Down's syndrome, dementia, multiple sclerosis, and amyotrophic lateral sclerosis. For diagnostic purposes, the expression of the protein of the invention could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals. For prevention and/or treatment purposes, the protein of the invention may be used to enhance electron transport and increase energy delivery using any of the gene therapy methods described herein or known to those skilled in the art.

In another embodiment, The invention further relates to methods and compositions using the protein of the invention or part thereof to diagnose, prevent and/or treat several disorders in which mitochondrial respiratory electron transport chain needs to be impaired, including but not limited to Sjogren's syndrome, Addison's disease, bronchitis, dermatomyositis, polymyositis, glomerulonephritis, diabetes mellitus, emphysema, Graves' disease, atrophic gastritis, lupus erythematosus, myasthenia gravis, multiple sclerosis, autoimmune thyroiditis, ulcerative colitis, anemia, pancreatitis, scleroderma, rheumatoid and osteoarthritis, asthma, allergic rhinitis, atopic dermatitis, dermatomyositis, polymyositis, and gout, using any techniques known to those skilled in the art including the antisense or triple helices strategies described herein.

Moreover, antibodies to the protein of the invention or part thereof may be used for detection of mitochondria organelles and/or mitochondrial membranes using any techniques known to those skilled in the art.

Protein of SEQ ID NO: 271 (internal designation 117-001-5-0-G3-CS)

The protein of SEQ ID NO: 271 is homologous to the family of lipopolysaccharide (LPS) binding proteins (LBPs). Several families of proteins have the ability to bind LPS including (a) the lipopolysaccharide-binding proteins (LBPs), and (b) the bactericidal permeability-increasing proteins (BPIs). Cholesteryl ester transfer protein (CETP), which is involved in the transfer of insoluble cholesteryl esters in reverse cholesterol transport, shares some homology to members of the LPS binding family of proteins.

Lipopolysaccharide (LPS), alternatively known as bacterial endotoxin, is a major component of the outer membrane of Gram-negative bacteria. It consists of serotype-specific O-side chain polysaccharides linked to a core oligosaccharide and Lipid A. LPS is a potent mediator of the inflammatory response and stimulates the expression of many pro-inflammatory and pro-coagulant compounds in monocytes, macrophages, and endothelial cells. While these responses are important in containing and eliminating localized infections, systemic exposure to LPS can lead to a number of

adverse effects. These include: (a) induction of an inflammatory cascade, (b) damage to the endothelium, (c) widespread coagulopathies, and (d) organ damage.

Systemic exposure to LPS can arise from direct infection by Gram-negative bacteria, leading to the complications of Gram-negative sepsis. Examples of diseases which are associated with Gram-negative bacterial infection or endotoxemia (including bacterial meningitis, neonatal sepsis, cystic fibrosis, inflammatory bowel disease, and liver cirrhosis), Gram-negative pneumonia, Gram-negative abdominal abscess, hemorrhagic shock, and disseminated intravascular coagulation. Subjects who are leukopenic or neutropenic, including subjects treated with chemotherapy or immunocompromised subjects, are particularly susceptible to bacterial infection and the subsequent effects of endotoxin exposure.

Gram-negative sepsis remains one of the primary causes of severe systemic inflammation in hospitalized and immunocompromised patients. Alternatively, changes in gut permeability by a variety of circumstances, including trauma, can lead to translocation of bacteria/LPS into the bloodstream. Bacteria translocated from the gut is thought to play a major role in post-surgical immunosuppression (Little et al., Surgery. 114: 87-91 (1993)) and hemorrhagic shock. Therefore, there is a great interest to characterize proteins involved in the biological response to LPS and to discover therapies that can counteract the effects of LPS in pathological situations.

LBP is a 60 kDa glycoprotein synthesized in the liver and present in normal human serum. LBP expression is upregulated in response to infectious, inflammatory, and toxic mediators. LBP expression has been induced in animals challenged with LPS, silver nitrate, turpentine, and *Corynebacterium parvum* (Geller et al., Surgery 128:22-28 (1993); Gallay et al., Infect. Immun. 61:378-383 (1993); Tobias et al., J. Exp. Med. 164:77-793 (1986)). LBP levels are correlated with exposure to LPS, and elevated levels (particularly persistent elevated levels) have been correlated with poor clinical outcomes in septic patients (U.S. Patent Nos. 5,484,705, and 5,804,367, hereby incorporated by reference in their entirety).

A portion of the LBP molecule (the N-terminal 1-197 aa) binds to the lipid A portion of the LPS molecule to form a high affinity LBP/LPS complex (Tobias, et al., J. Biol. Chem 264: 10867-10871 (1989)). The LBP/LPS complex potentiates the cellular response to LPS via an interaction with the monocytic differentiation antigen CD14 (Wright et al., Science. 249: 1431-1433 (1990); Lee et al., J. Exp. Med. 175:1697-1705 (1992)). LPS can be transferred from LBP to membrane-bound or soluble CD14. Activated CD14 can then interact with endothelial cells to elicit an inflammatory response. The C-terminal portion of LBP is required to transfer LPS to CD14 (U.S. Pat. No. 5,731,415; Theofan et al., J. Immunol. 152:3624-29 (1994); Han et al., J. Biol. Chem. 269:8172-75 (1994)). Evidence also suggests that LBP can neutralize LPS by an interaction with serum lipoproteins or through the internalization of an LBP/LPS/CD14 complex by neutrophils (Wurfel et al., J. Exp. Med. 180:1025-1035 (1994); Wurfel et al., J. Exp. Med. 181:1743-54 (1995); Gegner et al., J. Biol. Chem. 20:5320-5325 (1995)).

The subject invention provides the polypeptide of SEQ ID NO: 271 and polynucleotide sequences encoding the amino acid sequence of SEQ ID NO: 271. In a one embodiment, the polypeptides of SEQ ID NO: 271 are interchanged with the polypeptides encoded by the human cDNA of clone 181-20-3-0-B5-CS. Also included in the invention are biologically active fragments of the protein of SEQ ID NO: 271 and polynucleotide sequences encoding these biologically active fragments. In a preferred embodiment, biologically active fragments of SEQ ID NO: 271 are encoded by clone 181-20-3-0-B5-CS and comprise the first 181 amino acids encoded by clone 181-20-3-0-B5-CS. "Biologically active fragments" are defined as those peptide or polypeptide fragments of SEQ ID NO: 271 which have at least one of the biological functions of the full length protein (e.g., the ability to bind bacterial LPS).

The invention also provides variants of SEQ ID NO: 271. These variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence of SEQ ID NO: 271. Variants according to the subject invention also have at least one functional or structural characteristic of SEQ ID NO: 271, such as the biological functions described above. The invention also provides biologically active fragments of the variant proteins. Unless otherwise indicated, the methods disclosed herein can be practiced utilizing the polypeptide of SEQ ID NO: 271 or variants thereof. Likewise, the methods of the subject invention can be practiced using biological fragments of the protein of SEQ ID NO: or variants of said biologically active fragments.

Because of the redundancy of the genetic code, a variety of different DNA sequences can encode SEQ ID NO: 271. It is well within the skill of a person trained in the art to create these alternative DNA sequences which encode proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are, thus, within the scope of the subject invention. As used herein, reference to "essentially the same sequence" refers to sequences that have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity. Fragments retaining one or more characteristic biological activity of SEQ ID NO: are also included in this definition.

"Recombinant nucleotide variants" are alternate polynucleotides which encode a particular protein. They can be synthesized, for example, by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce specific restriction sites or codon usage-specific mutations, can be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic host system, respectively.

The protein of SEQ ID NO: 271, and variants thereof, can be used to produce antibodies according to methods well known in the art. The antibodies can be monoclonal or polyclonal. Antibodies can also be synthesized against fragments of SEQ ID NO: 271, as well as variants thereof, according to known methods. The subject invention also provides antibodies which

specifically bind to biologically active fragments of SEQ ID NO: 271 or biologically active fragments of SEQ ID NO: 271 variants.

The subject invention also provides for immunoassays which are used to screen for, monitor, or diagnose exposure to LPS. In one embodiment, diagnostic assays measure the level of LBP in patient plasma samples. LBP levels are known to rise in response to exposure to LPS, thus the measurement of the level of the protein of SEQ ID NO: 271 can provide an early indication of Gram-negative infection or of endotoxin exposure.

The subject invention provides methods of treating individuals infected with Gram negative bacteria comprising the administration of therapeutically-effective compositions comprising SEQ ID NO: 271. In one embodiment, the protein lacks the C-terminal portion (or portions of the C-terminal domain) necessary to transfer LPS to CD14. LPS can be scavenged by the excess N-terminal fragment and would be unable to induce an inflammatory response (see, *e.g.*, U.S. Patent No. 5,731,415, hereby incorporated by reference in its entirety).

Another aspect of the subject invention provides methods of prophylaxis. The method treats individuals by administration of therapeutically-effective amounts of compositions comprising SEQ ID NO: 271. Instances where this aspect of the invention can be performed include, but are not limited to, conditions associated with increased translocation of gut bacteria and endotoxin, particularly prior to surgery. In addition, patients who are at risk for potential Gram-infection, including but not limited to patients undergoing chemotherapy, or patients who are immunocompromised (for example with AIDS) can benefit from such treatment. Such uses are described in U.S. Patent No. 5,990,082, hereby incorporated by reference in its entirety.

The N-terminal portion of LBP, which lacks the ability to induce an inflammatory response, can be fused to other proteins or fragments thereof (such as the bactericidal/permeability-increasing protein or BPI) which can increase the association of these molecules with LPS and aid in the clearance of endotoxin from patients who have been exposed to Gram negative bacteria. Such preparations can be used to treat and inhibit a number of Gram-negative infections, Gram positive, or fungal infections, as described in the following patents: WO 95/19179 A, WO 95/19180 A, WO 95/19372 A, and WO 96/34873 A, each of which is incorporated by reference in its entirety.

The subject invention also provides methods of removing endotoxin from recombinantly-produced proteins. In one embodiment, the recombinantly-produced proteins are obtained from Gram-negative bacteria. In a preferred embodiment, the bacteria are *E. coli*. In another embodiment, the protein of SEQ ID NO: 271, biologically active fragments thereof, variants, or derivatives thereof, are contacted with compositions comprising recombinantly-produced proteins. The contacting step can take place with SEQ ID NO: 271 immobilized on a substrate or with SEQ ID NO: 271 present in free solution.

In addition, protein of SEQ ID NO: 271, biologically active fragments, or derivatives thereof, can be used in diagnostic assays to measure the level of LPS in patient plasma samples. In

such an assay, serum samples would be bound to a solid matrix, such as a membrane, plastic, treated plastic, or other supports, and then cloned with the protein of SEQ ID NO: 271. Visualization can be achieved by fusing protein of SEQ ID NO: to any number of enzymes followed by treatment with a chromogenic, fluorogenic, or luminescent substrate. Alternatively, the protein of SEQ ID NO: 271, biologically active fragments, variants, or derivatives thereof, can be linked to a fluorescent or luminescent protein or compound. The linkage can be chemical or made by recombinant techniques known to those skilled in the art. In addition, antibodies raised against the protein of SEQ ID NO: 271, biologically active fragments, variants, or derivatives thereof can be used to visualize the LPS/protein 271 complexes using immunoassays known to those skilled in the art.

Protein of SEQ ID NO:266 (internal designation 116-110-2-0-F4-CS)

The protein of SEQ ID NO:266, highly expressed in the testis, is encoded by cDNA of SEQ ID NO:25 and exhibits homology to the Ly-6 family of GPI-linked cell-surface glycoproteins composed of one or more copies of a conserved domain of about 100 amino-acid residues (PS00983; LY6_UPAR).

The protein of SEQ ID NO:266 shows significant structural similarities to mouse Ly-6 antigens, human CD59 and a herpes virus CD59 homolog. The protein of SEQ ID NO:266 displays one copy of the motif of the u-PAR/Ly-6 domain, with all ten extracellular cysteine residues conserved. The mature protein sequence contains a relatively high proportion of cysteine residues (10/105), which suggests that numerous disulfide bonds stabilize its tertiary structure. Furthermore, the 124 amino-acid long protein of SEQ ID NO:266 has a size very similar to that of many members of the Ly-6 family. In addition, the protein of the invention has a predicted signal peptide structure (positions from 1 to 19) and a C-terminal hydrophobic fragment (positions from 101 to 121) necessary for GPI-anchoring in a membrane. Thus, the protein of the invention has a clear evolutionary relationship with the Ly-6/uPAR family, particularly with Ly-6 subfamily.

The Ly-6/uPAR protein family members share one or several repeat units of the Ly-6/uPAR domain, which is defined by a distinct disulfide bonding pattern between 8 or 10 cysteine residues. This family can be divided into two subfamilies. One comprises GPI-anchored glycoprotein receptors with 10 cysteine residues. Another subfamily includes the secreted single-domain snake and frog cytotoxins, and differs significantly in that its members generally possess only eight cysteines and no GPI-anchoring signal sequence (Andermann K, et al. Protein Sci 8(4):810-819 (1999)). The Ly-6 family members are low molecular weight phosphatidyl inositol anchored glycoproteins with remarkable amino acid homology throughout a distinctive cysteine rich protein domain that is associated predominantly with O-linked carbohydrate. Their GPI links are necessary to anchor these cell surface proteins to the outside of the lipid bilayer membrane. The Ly-6 family includes human CD59, which protects from complement-mediated membrane damage, squid Sgp1

and Sgp2, urokinase plasminogen activator receptor, murine Sca-1 and Sca-2, and many other proteins. The general structure seen within the Ly-6 family resembles that of the receptor for a urokinase-type plasminogen activator and the alpha- neurotoxins from snake venoms (Fleming T J et al J Immunol 150:5379-5390 (1993); Ploug M and V Ellis FEBS Lett 349:163-168 (1994)).

The Ly-6 cell surface proteins are differentially expressed in several hematopoietic lineages that appear to function in signal transduction and cell activation predominantly on lymphoid cells in the mouse. Analyses using anti-Ly-6A/E monoclonal antibodies has also demonstrated in situ expression of Ly-6 molecules in brain tissue (staining primary associated with vascular elements throughout the brain). These proteins do not appear to be expressed during embryonic or neonatal stages of development (Cray C et al. Brain Res Mol Brain Res 8(1):9-15 (1990)).

Ly-6 protein expression has been shown to be factor-dependent. For example, the expression of the Ly-6A/E, which normally occurs in hemopoietic stem cells, fibroblasts, and T and B lymphocytes, has been shown to be greatly induced by IFN-ß in various tissues and cell lines. In addition, the Ly-6E Ag is associated with tyrosine kinases in T cells, and reduced expression of Ly-6E in T cells impairs normal functional responses, as well as tyrosine kinase activity, in these cells. Further, the IFNs are important in the generation of memory CD8+ T cells, and it has been demonstrated that the expression of Ly-6C Ag is a strong marker for the memory phenotype (Mehran M. et al. Journal of Immunology 163: 811-819 (1999)). Like their murine counterparts, a human homologue of Ly-6 genes, the 9804 gene, is responsive to IFNs. The 9804 gene is also inducible by retinoic acid during differentiation of acute promyelocytic leukemia cells. Further, cultured glial and neuronal cells express high levels of Ly-6A/E following incubation with cytokines, including rIFN-gamma. (Cray C et al. Brain Res Mol Brain Res 8(1):9-15 (1990)). Another member of the Ly-6 family, human protein RoBo-1, shows increased expression in response to two modulators of bone metabolism, estradiol and intermittent mechanical loading, suggesting a role in bone homeostasis (Noel LS et al. J Biol Chem, Vol. 273(7): 3878-3883 (1998)). Such factor-dependence of expression makes Ly-6 proteins either candidates or targets for alloresponses and autoimmune disease. For example, the high level factor-induced expression of LY-6s has been associated with lupus nephritis (Blake P G et al. J Am Soc Nephrol 4:1140-1150 (1993)).

Murine Ly-6 molecules have interesting patterns of tissue expression during haematopoiesis, from multipotential stem cells to lineage committed precursor cells, and on specific leukocyte subpopulations in the peripheral lymphoid tissues. These patterns suggest an intimate association between the regulation of Ly-6 expression, and the development and homeostasis of the immune system (Gumley TP et al. Immunol Cell Biol 73(4):277-296 (1995)). Ly-6M messenger RNA (mRNA) is easily detectable in hematopoietic tissue (bone marrow, spleen, thymus, peritoneal macrophages) as well as kidney and lung (Patterson JM et al. Blood 95(10):3125-3132 (2000)).

Normally, human blood cells are protected against autologous complement activation by membrane proteins that block the assembly of functional complement pores. One such protein is human Ly-6 CD59. Administration of CD59 prevents hemolytic disease or thrombosis. Further, the CD59 protein may prevent the complement-mediated lysis and activation of endothelial cells that leads to hyper acute rejection, and therefore may be administered during xenogeneic organ transplantation (Binette, J. P. and Binette, M. B., Scanning Microcs., 7:1107-10 (1993)).

The surface receptor for urokinase plasminogen activator (uPAR) has been recognized in recent years as a key molecule in regulating plasminogen mediated extracellular proteolysis. Surface plasminogen activation controls the connections between cells, basement membrane and extracellular matrix, and therefore the capacity of cells to migrate and invade neighboring tissues (Roldan AL et al. EMBO J 9(2):467-474 (1990)). Certain factors of the PA system, such as u-PAR, have been detected in organs of the male reproductive tract in various species. The morphological study provide support for the involvement of the PA system in human male reproductive physiology (Gunnarsson M et al. Mol Hum Reprod 5(10):934-940 (1999)).

LY-6 proteins have been suggested to play important roles in disorders such as cancers, nephopathies, autoimmune diseases, hemolytic disease, thrombosis, Alzheimer's disease, etc. Several members of the murine Ly-6 supergene family are clearly involved in the progression of certain mouse tumors, as their expression level is higher in highly malignant cells than in tumor cells with a lower malignancy phenotype. Sorting by flow cytometry of tumor cells to subpopulations expressing either high or low levels of Ly-6E.1 yielded cells expressing a high or a low malignancy phenotype, respectively. Further, it was shown that LY-6 is highly expressed on non-lymphoid tumor cells originating from a variety of tissues in mice. Upregulation or high expression is correlated with a more malignant phenotype which results in higher efficiency of local tumor production (Katz et al Int J Cancer 59:684-91 (1994)) .

Cells derived from angiogenic tumors express a higher tumorigenicity phenotype and a higher capacity to produce artificial pulmonary metastases than cells from the poorly angiogenic tumors. These cells also express significantly higher levels of the lymphocyte activation protein Ly-6E, so the angiogenic phenotype appears to be coregulated with Ly-6 (Sagi-Assif O et al. Immunol Lett 54(2-3):207-13 (1996)). Some LY-6 proteins also block secretion of interleukin II (IL-2) which is an approved anticancer agent and a key regulatory hormone in cell-mediated immunity (Fleming T J and T R Malek J Immunol 153:1955-62 (1994)). IL-2 stimulates the proliferation of both T and natural killer cells and activates NK cells which can directly lyse freshly isolated, solid tumor cells.

The high malignancy, high Ly-6E.1-expressing cells also expressed high levels of the receptor for urokinase plasminogen activator (uPAR), whereas low malignancy, low Ly-6E.1-expressing cells also expressed low levels of uPAR. Transfection studies have indicated that uPAR is causally involved in conferring a high malignancy phenotype upon tumor cells expressing high

levels of Ly-6E.1. E48, a human homologue of the murine ThB Ly-6 protein, is expressed on head and neck squamous carcinoma cells. In E48-stimulated cells, the binding of E48 to its microenvironmental ligand appears to transduce a signal that up-regulates the expression of the FX enzyme in these cells, leading to an increase in the levels of GDP-L-fucose (Rinat Eshel et al. J Biol Chem, Vol. 275(17):12833-12840 (2000)). A congenital disorder of leukocyte adhesion to vascular endothelium termed LADII is reflected in a generalized fucose deficiency and major defects in leukocyte trafficking and function. Ly-6 loss-variants of a murine tumor exhibit alterations in the incorporation of fucose and mannose into cellular glycoconjugates (Witz IP J. Cell. Biochem. Suppl. 34:61-66 (2000)).

It is believed that the protein of SEQ ID NO:266 is a novel member of the Ly-6 protein family, and is thus a specific cell-surface glycoprotein antigen involved in signal transduction and cell activation, proliferation and differentiation. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:266 from position 1 to position 18 and from position 19 to position 124. Other preferred polypeptides of the invention are any fragments of SEQ ID NO:266 having any of the biological activities described herein.

In one embodiment, this invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably testis. For example, the protein of the invention or part may be used to synthesize specific antibodies using any technique known to those skilled in the art. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, such as forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, etc., or to differentiate different tissue types in a tissue cross-section using immunochemistry.

Another embodiment of the present invention relates to methods of using of the protein of the invention or part thereof and related compounds and derivatives to diagnose developmental and malignant disorders in tissues including urogenital tissues and other tissues of the reproduction system of both sexes. For example, a biological sample is obtained from a patient with cancer or at risk of developing cancer, and the level of SEQ ID NO:25 polynucleotides or encoded polypeptides is detected within the cells of the sample. The detection of an elevated level of the SEQ ID NO:25 polynucleotides or encoded polypeptides in the sample relative to a control level indicates the presence of malignant cells within the patient. The expression of the protein of the invention can be investigated using any of a number of methods, including, but not limited to, Northern blotting, RT-PCR or immunoblotting.

Another embodiment of the invention relates to compositions and methods using the protein of the invention or part thereof in recombinant protein form as pharmacological agents in the treatment of developmental and malignant disorders in tissues including urogenital tissues and in other tissues of human reproduction system. Particulary, the protein of the invention or part thereof can be used in the treatment of disorders which are manifested by male sterility.

In another embodiment of the invention, antibodies which bind to the protein of the invention or part thereof are used in the treatment of tumors, e.g., human urogenital tumors, especially to enhance the secretion of interleukin II, which is an approved anticancer agent and key regulatory hormone in cell-mediated immunity. Such antibodies can be used alone or bound to a substance capable of ablating or killing cells as a therapy for urogenital disorders or cancers in which the protein of the invention is overexpressed.

The protein of the invention or part thereof may also be used in the treatment of diseases which can require transplantation, including various forms of cancers such as genitourinary cancers, carcinomas, sarcomas, atherosclerosis, angiogenesis, and benign tumors. As mentioned above, Ly-6 family includes several proteins which are similar to the protein of the invention and which are capable of protecting cells from complement-mediated membrane damage. Therefore, in another embodiment of the invention, recombinant proteins encoded by SEQ ID NO:25 or fragments thereof are administered during xenogeneic tissue transplantation to prevent complement-mediated lysis and to block activation of endothelial cells, which normally leads to hyper-acute rejection.

In addition, prevention of complement-mediated lysis may be particulary important in human and animal reproductive therapy, where functional survival of the germ cells during in vitro handling is crucial. Storage of sperm is of widespread importance in commercial animal breeding programs, human sperm donor programs, and in the treatment of certain disease states. For example, sperm samples may be frozen for men who have been diagnosed with cancer or other diseases that may eventually interfere with sperm production, as well as for assisted reproduction purposes where sperm may be stored for use at other locations or times. The procedures utilized in such cases include: washing a sperm sample to separate out the sperm-rich fraction from non-sperm components of a sample such as seminal plasma or debris; further isolating the healthy, motile sperm from dead sperm or from white blood cells in an ejaculate; freezing or refrigerating of sperm for use at a later date or for shipping to females at differing locations; extending or diluting sperm for culture in diagnostic testing or for use in therapeutic interventions such as in vitro fertilization or intracytoplasmic sperm injection (Cohen et al. 12 : 994-1001 (1997)). Once sperm have been washed or isolated, they are then extended (or diluted) in culture or holding media for a variety of uses (sperm analysis, diagnostic tests, assisted reproduction). Each of these uses for extended or diluted sperm requires a somewhat different formulation of basal medium (see, for review, US Patent No. 6,140,121 Ellington et al. Oct. 2000); however, in all cases sperm survival is suboptimal outside of the female reproductive tract. Novel additional components of a dilution or storage medium which could improve the functional preservation of sperm would be useful. Therefore, in another preferred embodiment of this invention, purified recombinant proteins encoded by SEQ ID NO:25 or fragments thereof can be added as components of pharmacological media designed to protect spermatozoa. The methods used to compose such preservation media are generally known by those skilled in the art (for ex., Oliver S.A . et al. US patent 5,897,987 Apr.1999; Cohen J. et al.,

supra). Inversely, in yet another embodiment of this invention, ligands, inhibitors, neutralizing antibodies or other biological agents which recognize the protein of the invention and which bind it and which block it can be used as components of pharmacological formulations designed for male contraception purposes.

In still another embodiment of this invention, chimeric ligands or derivatives which recognize the protein of the invention or part thereof and which could be internalized into cell can be used to design a system of drug delivery finely targeted toward urogenital and other tissues which express the protein of SEQ ID NO:266. For example, such recognizing molecules can be incorporated into the membranes of liposomes to allow the specific delivery of the liposomes to cells expressing the protein of SEQ ID NO:266. Methods of designing such drug delivery systems are known by those skilled in the art (Smith H.J. Introduction to the principles of drug design and action, 3rd ed. (1998)).

Proteins SEQ ID NOs:417, 413, 418 (internal designations 188-45-1-0-D3-CS, 188-26-4-0-F5-CS, and 188-5-1-0-H6-CS)

The proteins of SEQ ID NOs:417, 413, and 418, encoded by the cDNAs of SEQ ID NOs: 176, 172, and 177, are expressed in the brain and exhibit strong homology with proteins with redox activity (see, e.g. Genbank accession numbers AK001293 and AF029689, and Geneseqp accession number: Y59180).

The protein of SEQ ID No:418 (320 amino acids) is a variant of AK001293 (322 amino acids). AK001293 has six extra nucleotides, within the same ORF, as SEQ ID No:418, producing a longer protein. SEQ ID NO:418 exhibits the Pfam Zinc-binding dehydrogenase (adh zinc) signature from positions 16 to 313. SEQ ID NO:418 presents all the conserved residues of the motif except for a histidine that is thought to be a zinc-ligand. This lack of zinc-ligand residues is a feature of the quinone oxidoreductases (QOR), a subfamily of zinc-binding dehydrogenases.

SEQ ID NO:413 (191 amino acids) shares the first 172 amino acids with SEQ ID NO:418. The deletion of one nucleotide at position 583 in the SEQ ID NO:413 cDNA sequence (corresponding to amino acid 173), however, creates a change of ORF compared to SEQ ID NO:418 and AK001293.

SEQ ID NO:417 is a short protein (20 amino acids) whose sequence corresponds to the N-terminal end of the other proteins of the invention. The presence of a T (instead of a G in public sequences and SEQ ID NOs:413 and 418) at position 128 on the cDNA creates a STOP codon, creating a shorter protein.

SEQ ID NOs:417, 413 and 418 are similar to the QORs, a family of zinc-binding dehydrogenases. QORs are cytoplasmic redox-regulated flavoenzymes that catalyze the one or two-electron reduction of quinones. QORs bind NADP and are inhibited by dicoumarol.

The activity of QORs protects cells against toxicity, mutagenicity, and cancer due to exposure to environmental and synthetic quinones and their precursors. Thus, QORs play a central role in monitoring cellular redox state and act to protect against oxidative stress induced by a variety of metabolic situations (Raina A.K. et al. (1999) Redox Rep. 4:23-7). The oxidoreductase activity also permits the activation of bioreductive anticancer drugs (Begleiter A. et al. (1996) Br. J. Cancer Suppl. 27:S9-14).

The metabolism of quinones involves enzymatic reduction of the quinone by one or two electrons. In the activation of quinone-containing antitumor agents, this reduction results in the formation of the semiquinone or the hydroquinone of the anticancer drug. The consequence of these enzymatic reductions is that the semiquinone yields its extra electron to oxygen with the formation of superoxide radical anion and the original quinone. This reduction by a reductase followed by oxidation by molecular oxygen (dioxygen) is known as redox-cycling and continues until the system becomes anaerobic. In the case of a two-electron reduction, the hydroquinone could become stable, and as such, be excreted by the organism in a detoxification pathway.

The cellular antioxidant response is mediated by a battery of detoxifying/defensive proteins. The promoters of genes that encode these proteins contain a common cis-element termed the antioxidant response element (ARE). Many transcription factors, including Nrf, Jun, Fos, Fra, Maf, YABP, ARE-BP1, Ah (aromatic hydrocarbon) receptor, and estrogen receptor bind to the ARE from various genes. Among these factors, Nrf-Jun heterodimers positively regulate ARE-mediated expression and induction of genes in response to antioxidants and xenobiotics (reviewed in Dhakshinamoorthy S. et al. (2000) Curr. Top Cell Regul. 36:201-16). On the other hand, c-Fos represses ARE-mediated gene expression (Venugopal, R., and Jaiswal, A.K. (1996) Proc. Natl. Acad. Sci. USA 93, 14960-5).

Elevated levels of QOR activity have been reported in several kinds of tumors such as liver, colon, lung and breast (Belinsky M., Jaiswal A.K., (1993) Cancer Metastasis Rev 12:103-17). Bioreactive antitumor agents are an important class of anticancer drugs that require activation by reduction. For this reason, QORs are a potential target on which to base the development of new antitumor compounds. Certain QORs have already been implicated in the metabolism, activation and mechanism of cytotoxicity of some anticancer drugs such as mitomycin C, indoloquinone E09 (Ross D. et al. (1994) Oncol. Res. 6:493-500), CB 1954 (Knox R.J. et al. (2000) Cancer Res. 60:4179-86) or antiestrogens in breast cancer (Montano M.M., Katzenellenbogen B.S. (1997) PNAS 94:2581-6).

In addition, some of the proteins of the QOR family are thought to play a role in the prevention of apoptosis following oxidative stress. The tumor suppressor gene p53 has been directly implicated in the induction of apoptosis in dividing cells and in hippocampal pyramidal neurons (Jordan J. et al. (1997) J. Neurosci 17:1397-405) and a QOR gene has been described as a p53-regulated gene (Kostic C, Shaw P.H. (2000) Oncogene 19:3978-87).

It is believed that the proteins of SEQ ID NOs:417, 413 and 418 have a redox activity, most likely as QORs. Thus, they are expected to act as an endogenous antioxidant against oxidative stress and may be able to use NADP as cofactor. The proteins of the invention may be used to deactivate toxins and to activate bioreductive anticancer drugs. In addition, they may prevent apoptosis following oxidative stress and be regulated by p53. Because proteins SEQ ID NOs:417 and 413 do not contain the Pfam Zinc-binding dehydrogenase (adh zinc) signature, in contrast to SEQ ID NO:418, they may act as a competitive inhibitor, i.e. a dominant negative form, of the functional protein.

The oxidoreductase activity of the proteins of the invention may be assayed using any technique known to those skilled in the art. For example, the measurement of the rate of oxidation of NADPH and oxygen consumption, and the detection of the semiquinone and reactive oxygen species, may be performed as described by Gutierrez P.L. (Gutierrez P.L . (2000) Front. Biosci. 5:D629-38), or by any other method skilled in the art. The enzymatic activity of the proteins of the invention in different affected and control tissues may be assayed by histochemical staining. To confirm the role of the proteins of the invention in the cellular antioxidant response, in vitro and in vivo assays may be performed. Transcription levels of the genes coding for the proteins of the invention may be measured using standard techniques after exposure to quinones or derived compounds as beta-naphtoflavone (beta-NF), as described by Belinsky M. and Jaiswal A.K. (supra), as well as in response to transcription factors such as Nerf, Jun and c-Fos, or in the presence of p53.

In one embodiment of the present invention, the present protein can be used to detect specific cell types in vitro or in vivo. For example, as the present proteins are overexpressed in the brain, reagents capable of specifically recognizing the present protein can be used as markers for brain cells. Brain-specific markers have a number of uses, including for the identification of specific tissues for histological analyses, as well as to detect the origin of tumor cells. In addition, as the expression of the present protein is likely induced by transcription factors such as Nrf, Jun, Fos, Fra, Maf, YABP, ARE-BP1, Ah (aromatic hydrocarbon) receptor, and estrogen receptor, as well as by p53, reagents specific for detecting the present protein can also be used as a marker for the activity of any of these proteins in vitro or in vivo. In view of the association between many of these proteins and diseases such as cancer, the ability to detect the presence or absence of the proteins provides powerful tools for disease diagnosis and screening. For any of these applications, the expression of the present protein can be detected using any standard method, including Northern blots, western blots, in situ hybridization, PCR, etc.

In another embodiment, the proteins of the invention can serve as markers for cellular oxidative stress in vivo and in vitro. As such, the proteins of the invention or part thereof may be useful in the diagnosis of disorders in which oxidative stress is implicated, including a large variety of types of cancer as well as neurodegenerative disorders such as Alzheimer's disease (AD), amyothropic lateral sclerosis (ALS) or Parkinson disease (PD). For diagnostic purposes, the

expression of the protein of the invention may be investigated using, e.g. Northern blotting, RT-PCR or immunoblotting methods and compared to the expression in control individuals. An increased levels of the proteins of the invention in patients compared with controls indicates a major shift in redox balance and, thus, indicates the presence of the disease or of a susceptibility for the disease.

The invention further relates to methods and compositions using the proteins of the invention or part thereof to prevent and/or treat disorders in which oxidative stress is implicated, including those mentioned above. For these purposes the proteins themselves, or polynucleotides encoding the proteins, or an activator of protein expression may be administrated to patients, or to disease-free individuals in case of increased susceptibility to one of these disorders.

In another embodiment, the protein of the invention or part thereof is used to prevent cells from undergoing apoptosis. They may thus be useful in the diagnosis, treatment and/or prevention of disorders and processes in which apoptosis is deleterious, including but not limited to immune deficiency syndromes (including AIDS), type I diabetes, pathogenic infections, cardiovascular and neurological injury, alopecia, aging, degenerative diseases including AD and PD, dystonia, Leber's hereditary optic neuropathy and schizophrenia. For all such diagnostic purposes, the expression of the proteins of the invention can be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals.

The invention relates to methods and compositions using the proteins of the invention or part thereof as detoxifying enzymes against quinones. There are a variety of quinones with a toxic effect in cells (e.g. quinones derived from the oxidation of phenolic metabolites of benzene, DA-quinones, or menadione). Thus, the proteins of the invention or part thereof may be protective against the hematotoxic and carcinogenic effects of benzene, as well as against benzene-caused diseases such as cancer, aplastic anemia and pancytopenia. Moreover, they may detoxify DA-quinones in the brain, thereby providing neuroprotection in Parkinson's Disease. In still another embodiment, the proteins of the invention or part thereof may protect cells against menadione-induced oxidative stress, with known effects on myocardial cells (Floreani M. et al (2000) Biochem Pharmacol. 60:601-5). For prevention and/or treatment purposes the proteins themselves, or polynucleotides encoding the proteins, or an activator of protein expression may be administrated.

In another embodiment, the present proteins may be a target of chemotherapy specific to different kinds of cancer, to ensure a favorable response to anticancer drugs. Specifically, proteins of the invention or part thereof may be used as an activator of cytotoxic prodrugs of quinone family. Accordingly, the protein of the invention or part thereof may be administered to a patient in conjunction with a bioreductive anticancer agent in order to activate the drug. This co-administration may be by simultaneous administration, such as a mixture of the oxidoreductase and the drug, or by separate simultaneous or sequential administration. Cancer-specific antitumor agents based on QOR substrates may be designed as described by Xing J. et al. (Xing J. et al. (2000)

Med. Chem. 43:457-66) and assayed as described in Li B. et al. (Li B et al. (1999) Chem. Res. Toxicol. 12:1042-9). Alternatively, as the present proteins may be overexpressed in tumor cells, such methods may be performed by simply detecting the level of the present protein in tumor cells, and administering the prodrug specifically to those patients found to have elevated levels of the protein in their tumor cells.

Proteins of SEQ ID NOs: 415, 310, 317 (internal designation 188-29-2-0-H1-CS, 188-18-4-0-A9-CS, 188-9-2-0-E1-CS)

Mammalian inositol hexakiphosphate kinase 2 (IP6K2), an enzyme of the inositol phosphate pathway, has been cloned and described by two independent groups [Saiardi, A.; Erdument-Bromage, H.; Snowman, A. M.; Tempst, P.; and Snyder, S. H., (1999) Current Biology 9, 1323-1326, and Katai, K.; Miyamoto, K-I.; Kishida, S.; Segawa, H.; Nii, T.; Tanaka, H.; Tani, Y.; Arai, H.; Tatsumi, S.; Morita, K.; Taketani, Y.; and Takeda, E. (1999) Biochem. J. 343, 705-712]. Newly identified consensus sequences of inositol-polyphosphate kinases are represented by [LV]-[LA]-[DE]-X(3-8)-P-X-[VAI]-[ML]-D-X-K-[ML]G [Saiardi, A.; Erdument-Bromage, H.; Snowman, A. M.; Tempst, P.; and Snyder, S. H. (1999) Current Biology 9, 1323-1326]. IP6K2 catalyzes the transfer of phosphate groups from InsP6 or Ins(1,3,4,5,6)P5 (the substrate), to another protein or small molecule, such as a nucleoside di-phosphate.

The subject invention provides the polypeptides of SEQ ID NOs:415, 310, and 317, encoded by the cDNAs of SEQ ID NOs:174, 69, and 76, respectively. The invention also provides biologically active fragments of SEQ ID NOs:415, 310, and 317. In one embodiment, the polypeptides of SEQ ID NOs:415, 310, and 317 are interchanged with the corresponding polypeptides encoded by the human cDNA of clone 188-29-2-0-H1-CS, 188-18-4-0-A9-CS, or 188-9-2-0-E1-CS. "Biologically active fragments" are defined as those peptide or polypeptide fragments having at least one of the biological functions of the full length protein (e.g., kinase activity). Compositions of the protein/polypeptide of SEQ ID NOs:415, 310, or 317, or biologically active fragments thereof, are also provided by the subject invention. These compositions may be made according to methods well known in the art.

The invention also provides variants of the protein of SEQ ID NOs:415, 310, or 317. These variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequences encoded by SEQ ID NOs:415, 310, and 317. Variants according to the subject invention also have at least one functional or structural characteristic of the protein of SEQ ID NOs:415, 310, or 317. The invention also provides biologically active fragments of the variant proteins. Compositions of variants, or biologically active fragments thereof, are also provided by the subject invention. These compositions may be made according to methods well known in the art. Unless otherwise indicated, the methods disclosed herein can be practiced utilizing the protein encoded by SEQ ID

NO:415, 310, or 317, biologically active fragments of SEQ ID NO:415, 310, or 317, variants of SEQ ID NO:415, 310, or 317, and biologically active fragments of the variants.

Because of the redundancy of the genetic code, a variety of different DNA sequences can encode the amino acid sequence of SEQ ID NO:415, 310, or 317. In a preferred embodiment, SEQ ID NO:415, 310, or 317 is encoded by clone 188-29-2-0-H1-CS, 188-18-4-0-A9-CS, or 188-9-2-0-E1-CS, or by the cDNAs of SEQ ID NO:174, 69, or 76. It is well within the skill of a person trained in the art to create these alternative DNA sequences which encode proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are, thus, within the scope of the subject invention. As used herein, reference to "essentially the same" sequence refers to sequences that have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity. Fragments retaining one or more characteristic biological activity of the protein encoded by SEQ ID NO:415, 310, or 317 are also included in this definition.

In one aspect of the subject invention, SEQ ID NO:415, 310, or 317, and variants thereof, can be used to generate polyclonal or monoclonal antibodies. Both biologically active and immunogenic fragments of SEQ ID NO:415, 310, or 317, or variant proteins, can be used to produce antibodies. Polyclonal and/or monoclonal antibodies can be made according to methods well known to the skilled artisan. Antibodies produced in accordance with the subject invention can be used in a variety of detection assays known to those skilled in the art. The antibodies may be used to agonize or antagonize the biological activity of the protein of SEQ ID NO:415, 310, or 317.

The protein of SEQ ID NO:415, 310, or 317 can be used for the synthesis of nucleoside triphosphate (NTP) compounds. In one embodiment, the NTP compound produced is ATP, GTP, CTP, or TTP. In this aspect of the subject invention, SEQ ID NO:415, 310, or 317 removes a phosphate from InsP6 or Ins(1,3,4,5,6)P5 and transfers it to a nucleoside diphosphate (e.g., ADP, CTP, GDP, or TDP) to create a NTP. The conditions and methods for the synthesis of NTP compounds, such as ATP, are well known to the skilled artisan. Thus, the protein of SEQ ID NO:415, 310, or 317 has industrially useful function for the synthesis of commercially valuable products.

The subject invention also provides methods of determining the relative amounts of InsP6 or Ins(1,3,4,5,6)P5 in the cell by a kinase assay. In this aspect of the invention, SEQ ID NO:415, 310, or 317 can be used to transfer phospate groups from InsP6 or Ins(1,3,4,5,6)P5 to acceptor substrates according to well-known kinase activity assays.

Protein of SEQ ID NO:294 (internal designation 181-16-2-0-A7-CS)

The protein of SEQ ID NO:294 is encoded by the cDNA of SEQ ID NO:53. It will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:294 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 181-16-2-0-A7-CS. In addition, it will be appreciated that all characteristics and uses of the

nucleic acid of SEQ ID NO:53 described throughout the present application also pertain to the human cDNA of clone 181-16-2-0-A7-CS.

This gene was isolated from fetal liver and expression has also been detected in fetal kidney, placenta, liver, brain, hypertrophic prostate, salivary gland and testis. Data from PCT application WO 98/23435 indicate expression is primarily in bone marrow cell lines, and to a lesser extent, in human endometrial stromal cells, human adult small intestine and human pancreas tumor. PCT application WO 99/14484 reports the fraction of expression in the gastrointestinal system (0.227), reproductive system (0.193), and hematopoietic/immune system (0.168). Finally, this protein is 55% identical and 76% similar to CGI-128 protein, which was isolated from CD34+ cells and is also found in cell lines from the hematopoietic lineage including, HL6 (granulocytic), Jurkat (T-lymophocytic), K562 (erythro-megakaryocytic), and U937 (monocytic).

Supernatant harvested from cells expressing the product of this gene has been shown to increase the permeability of the plasma membrane of renal mesangial cells to calcium. Thus, it is believed that the product of this gene is involved in activating a signal transduction pathway when it binds a receptor on the surface of the plasma membrane of both mesangial cells and other cell types, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses, which include, but are not limited to, activating mesangial cells by contacting said cells with a full length polypeptide or a polypeptide fragment which demonstrates this biological activity. Further, the polynucleotides and polypeptides can be used in the methods described in WO9915652, incorporated in its entirety. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane potential. Alterations in small molecule concentration can be measured to identify supernatants, which bind to receptors of a particular cell. In addition, when tested against fibroblast cell lines, supernatants removed from cells containing this gene activated the EGR1 (early growth response gene 1) promoter element. Thus, it is likely that this gene activates fibroblast cells through the EGR1 signal transduction pathway. EGR1 is a separate signal transduction pathway from Jak-STAT, genes containing the EGR1 promoter are induced in various tissues and cell types upon activation, leading the cells to undergo differentiation and proliferation (PCT application WO 98/23435)

Polynucleotide comprising sequences encoding the signal peptide of the protein, e.g. VLWLSGLSEPGAA/RQ, can be used in construction of secretion vectors. These vectors would then facilitate the secretion of fusion proteins into the media of cells that have been transfected with the construct of interest. Antibodies which specifically bine the signal peptide could be used to purifiy the fusion protein from the media if desired.

Polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, haemopoietic and gastrointestinal tract

disorders and stromatosis, in addition to endothelial, mucosal, or epithelial cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides, are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and digestive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g. hemaopoietic, immune, reproductive, gastrointestinal, endocrine, and cancerous and wounded tissues) or bodily fluids (e.g. lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distributioin in bone marrow cells, fetal liver and fetal kidney, combined with the detected calcium flux and EGR1 biological activity, indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune and gastrointestinal tract disorders, and stromatosis, particularly tumors and proliferative disorders. More specifically, polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The polypeptides and polynucleotides of the invention can be used to enhance hematopoesis as described in WO9831385, incorporated in its entirety. The uses include bone marrow cell ex vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Additionally, since the gene product of 181-16-2-0-A7-CS has been shown to activate the EGR1 promoter element, it likely activates EGR1 signaling activity in fibroblasts. Recent data shows that activation of EGR1 plays a role in wound repair. The cellular transcription factor early growth response factor 1 (Egr1) is expressed minutes after acute injury and serves to stimulate the production of a class of growth factors whose role is to promote tissue repair. Egr-1 expression at the site of dermal wounding in rodents promotes angiogenesis in vitro and in vivo, increases collagen production, and accelerates wound closure. These results show that Egr-1 gene therapy accelerates the normal healing process (Human Gene Ther 2000, vol 11(15):2143-58). Thus, an activator of EGR1 signaling, specifically the gene products of 181-16-2-0-A7-CS (polypeptides and polynucleotides), would be useful in the wound healing process using the methods described in WO9941282 and WO9932135, incorporated by reference in their entireties.

Protein of SEQ ID NO:305 (internal designation 187-37-0-0-c10-CS)

The protein of SEQ ID NO:305, encoded by the cDNA of SEQ ID NO:64, is highly expressed in the prostate and brain. The protein of the invention is strongly homologous to the D9 protein, found in both humans (GNP accession number: U95006 and U95007) and in mice (GNP accession number: U95003, U95004, and U95005). D9 is a myeloid precursor protein transcript regulated by the retinoic acid receptor α, hereafter referred to as RAR-α (Scott et al. Blood 1996; 88: 2517-30).

Retinoic acid is the active metabolite of vitamin A, which contributes to a wide range of biological processes such as cellular differentiation, embryogenesis, and tumor suppression. More specifically, retinoic acid stimulates myeloid precursor differentiation into mature granulocytes. For instance, *in vitro* treatment of acute promyelocytic leukemia blast cells with retinoic acid induces their differentiation (Miyauchi et al. Leuk Lymphoma 1999;33:267-80). In addition, treatment with retinoic acid can induce disease remission in patients affected with promyelocytic leukemia by causing granulocyte precursor differentiation (Slack et al. Ann Hematol 2000;79:227-38).

The diverse range of responses to retinoic acid are mediated by three receptor subtypes: RAR-α, RAR-β, and RAR-γ. RAR-α has been identified as being important for bone marrow maturation of granulocytes (Tsai et al. Genes Dev 1992;6:2258-69). In addition, RAR-α is almost invariably involved in acute promyelocytic leukemia cells by a reciprocal translocation between the long arms of chromosomes 15 and 17 (Alcalay et al., Proc Natl Acad Sci USA 1991;88:1977-81). This type of leukemia is mainly characterized by a predominance of malignant promyelocytes, and severe hemorragic manifestations resulting from activation of the coagulation cascade and the fibrinolytic system (Tallman et al. Semin Thromb Hemost 1999;25:209-15). Reciprocal chromosomal translocation leads to the production of a fusion protein that inhibits differentiation and promotes survival of myeloid precursor cells (Grignani et al. Cell 1993;74, 423-431). Transient transfection of a vector containing RAR-α in a promyelocyte cell line causes upregulation in an early manner of several genes, including D9, which is strongly related to protein of SEQ ID NO:305 (Scott et al. Blood 1996; 88: 2517-30). Thus, it is believed that the protein of SEQ ID NO:305 is a myeloid-related protein whose expression is induced by the activation of retinoic acid receptors, including RAR-α.

In a preferred embodiment, the protein of the invention or part thereof may be used to assay the activity of RAR-α protein or retinoic acid in a biological sample. Specifically, as the expression of the protein is believed to be under the direct control of retinoic acid receptors, the level of the protein of the invention, or of the mRNA encoding the protein, can serve as a sensitive and immediate marker for the effects of retinoic acid upon a cell. An ability to detect retinoic acid receptor activation in cells using the present protein has numerous uses. For instance, the protein of the invention or part thereof can be used to monitor the effects of retinoic acid on cells of a patient

undergoing retinoic acid treatment for promyelocytic leukemia (Slack et al. Ann Hematol 2000;79:227-38). As retinoic acid treatment is associated with frequent retarded dose-dependant side effects, it is believed that an assay based on protein of SEQ ID NO:305 could be used to adjust the dose of retinoic acid administered in patients affected with promyelocytic leukemia, in order to predict and avoid such adverse side-effects (Slack et al. Ann Hematol 2000;79:227-38).

In another embodiment, the present polypeptides and polynucleotides can be used to identify myeloid precursors, as well as brain and prostate tissues. The ability to specifically visualize myeloid precursor cells, as well as brain and prostate tissues (and cells derived from the tissues), is useful for any of a number of applications, including to determine the origin or identity of, e.g. cancerous cells, as well as to facilitate the identification of particular cells and tissues for, e.g. the evaluation of histological slides. In addition, such assays can be used to examine the extent of differentiation in myeloid precursor cells.

The present invention further relates to *in vitro* assays and diagnostic kits based on the protein of the present invention or part thereof. Such assays may be used for diagnosis of disorders where the protein activity is abnormally downregulated, such as cancer, and hematological disorders including leukemia. As the protein of SEQ ID NO:305, RAR-α, and acute promyelocytic leukemia are all related, variation in the measured level of the present protein of the invention or part thereof can be used as a diagnostic or screening test for acute promyelocytic leukemia, e.g. using a biological sample such as serum or plasma. Further, an assay that can detect an abnormal level of the protein of the invention or part thereof can be used to detect residual disease in acute promyelocytic leukemia. Such an assay may be used to aid therapeutic decisions in this disorder, e.g. more or less aggressive treatments, the duration of treatment, etc.

In another embodiment, various methods can be used to modulate activity and/or expression of the protein of SEQ ID NO:305, e.g. for the treatment, attenuation and/or prevention of various disorders. In one embodiment, any of a number of reagents, e.g. polynucleotides encoding the protein of SEQ ID NO:305 or a fragment thereof, the protein of SEQ ID NO:305 itself, or a compound that increased the expression or activity of the protein of SEQ ID NO:305, can be administered to a patient suffering from, or at risk of developing, various disorders including cancer, and hematological diseases such as leukemia, and neutropenia. For instance, but not limited to it, proteins or other capable of enhancing the expression or activity of the protein of SEQ ID NO:305 can be administered to treat patients affected with acute promyelocytic leukemia, in order to induce differentiation of the affected cells into mature granulocytes (Slack et al. Ann Hematol 2000;79:227-38). In still another preferred embodiment, proteins or other compounds capable of increasing the expression or activity of the protein of the invention can be used to treat, prevent and/or attenuate neutropenia or agranulocytosis patients, in order to induce *in vivo* differentiation of myeloid precursors into mature granulocytes. In still another preferred embodiment, proteins or other compounds capable of increasing the expression or activity of the protein of SEQ ID NO:305

can be used to treat coagulopathic diseases, such as thrombosis or hemorragic manifestations. For instance, they can be used to treat disseminated intravascular coagulation, a severe hemorragic syndrome. This embodiment is supported by the fact that acute promyelocytic leukemia is frequently associated with disseminated intravascular coagulation (Tallman et al. Semin Thromb Hemost 1999;25:209-15), and disseminated intravascular coagulation is efficiently corrected with retinoic acid (Dombret et al. Leukemia 1993;7:2-9).

In addition, modulation of the expression or activity of the protein of the invention can be used to modulate differentiation of cells, e.g. promyelocytic leukemia. In one such embodiment, the protein of the invention is inhibited, e.g. using antisense molecules, antibodies, or small molecule inhibitors of the expression or activity of the protein, in order to maintain the undifferentiated state of cells grown in vitro. Alternatively, agents that increase the expression or activity of the protein in cells can be used to induce cellular differentiation, e.g. in the preparation of specific cell types in vitro for particular therapeutic applications.

Protein of SEQ ID NO:248 (internal designation 105-035-2-0-C6-CS) and SEQ ID NO:313 (internal designation 188-28-4-0-D4-CS)

The proteins of SEQ ID NO:248, encoded by the cDNA of SEQ ID NO:7, and SEQ ID NO: 313, encoded by the cDNA of SEQ ID NO:72, are highly expressed in brain, liver, pancreas, and testis. The proteins of the invention are nuclear proteins (Miller et al. J Biol Chem 2000;275:32052-6) that display a membrane-spanning segment from amino acids 58 to 78. These proteins are homologous to the human RNA polymerase II elongation factor ELL3 (EMBL accession number AF276512 ; Miller et al. J Biol Chem. 2000; 275:32052-6). In addition, the proteins of SEQ ID NO:248 and SEQ ID NO:313 share sequence homology with two other members of the polymerase II elongation factor family: ELL, and ELL2. The protein of SEQ ID NO:313 is similar to the N-terminal sequence the protein of SEQ ID NO:248, but differs after residue 240 because of a frameshift that produces a premature stop in the sequence SEQ ID NO:72 (Miller et al. J Biol Chem 2000; 275:32052-6). Additionally, the alignment of the protein of SEQ ID NO:248 with occludin, an integral membrane protein found at tight junctions (Furuse et al. J Cell Biol 1994; 127:1617-26), reveals that both proteins display a C-terminal ZO-1 binding domain, with a 26% homology over a 108 amino acid segment. Protein SEQ ID NO:313 lacks this domain, as its C-terminal region is truncated as compared to the protein of SEQ ID NO:248. ZO-1 is part of the family of membrane-associated guanylate kinase homologs (MAGUKs) believed to be important in signal transduction originating from sites of cell-cell contact (Willott et al. Proc Natl Acad Sci USA 1993; 90:7834-8).

The proteins of SEQ ID NOs:248 and 313 are RNA polymerase II elongation factors that increase the catalytic rate of transcription elongation, a phase during which RNA polymerase II moves along the DNA and extends the growing RNA chain (Miller et al. J Biol Chem 2000;

275:32052-6). Specifically, the proteins of SEQ ID NOs:248 and 313 suppress transient pausing at multiple sites along the DNA, thereby altering the K_m and/or the V_{max} of the polymerase (Miller et al. J Biol Chem 2000; 275:32052-6). The present proteins belong to a family that is known to include one virally encoded protein (Tat) and six cellular proteins (SIX, P-TEFb, TFIIF, Elongin (SIII), ELL and ELL2).

A growing body of evidence suggests that mis-regulation of elongation may be a key element in a variety of human diseases (see, Aso et al. J Clin Invest 1996; 97:1561-9). For instance, two RNA polymerase II elongation proteins have been implicated in oncogenesis: ELL, which is a frequent target for translocation in acute myeloid leukemia (Thirman et al. Proc Natl Acad Sci USA 1994; 91:12110-4 ; Mitani et al. Blood 1995;85:2017-24), and elongin, which is a transcription factor regulated by the product of the von Hippel-Lindau tumor suppressor gene, which is itself mutated in the majority of clear-cell renal carcinomas and in families with von Hippel-Lindau disease (Duan et al. Science 1995;269:1402-6, Kibel et al. Science 1995; 269:1444-6). In addition, overexpression of ELL leads to the transformation of fibroblasts (Kanda et al. J Biol Chem. 1998 27; 273:5248-52). Thus, the proteins of SEQ ID NOs:248 and 313 may be important for oncogenesis of multiple types of neoplastic diseases, especially hematological malignancies.

In one embodiment, the present proteins are used to increase the rate of transcription in vitro. Such an increase can be used for any of the large number of in vitro transcription reactions which are routinely used for molecular biological applications, e.g. for the preparation of RNA, for protein production, for the characterization of promoters and transcription factors, etc.

In another embodiment, the present invention provides diagnostic tools for the detection of mutations in the genes encoding SEQ ID NOs:248 or 313. Such mutations may be detected by a variety of techniques, including RNase and S1 protection assays; alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents such as SSCP or DGGE; dHPLC; and direct DNA sequencing. The detection of mutations in the genes encoding SEQ ID NOs:248 or 313 are useful for the detection of a number of diseases and conditions, such as cancers and hematological malignancies including leukemia. For example, the RNA polymerase II Elongation Factor ELL gene undergoes frequent translocations in acute myeloid leukemia (Thirman et al. Proc Natl Acad Sci USA 1994; 91:12110-4 ; Mitani et al. Blood 1995; 85:2017-24), and it is likely that other elongation factors are involved in additional such diseases.

Another embodiment of the present inventions relates to compositions and methods for using the proteins or part thereof to specifically visualize myeloid precursor cells, as well as pancreas, liver and testis tissues (and cells derived from the tissues). The ability to detect such cell types is useful for any of a number of applications, including to determine the origin or identity of, *e.g.* cancerous cells, as well as to facilitate the identification of particular cells and tissues for, *e.g.* the evaluation of histological slides. In addition, such methods can be used to examine the extent of differentiation in myeloid or myeloid-progenitor cells for staging of leukemia or any other

neoplastic disorder. Any method for detecting the presence of the proteins of the invention, or nucleic acids encoding the proteins, can be used, including methods involving the use of antibodies immunospecific for the proteins of invention. Such antibodies can be used in various methods including radioimmunoassays, competitive binding assays, Western Blot analysis and enzyme-linked immunosorbant assay (ELISA) assays, or any other technique known to those skilled in the art. In another embodiment, the present protein or part thereof can be used for the treatment, attenuation and/or prevention of conditions associated with unbalanced amounts and/or activity of the protein of SEQ ID NO:248 or 313. Other modulatory substances can also be used in such embodiments, including chemical compounds such as agonists and antagonists, nucleic acids including antisense and ribozyme sequences, and antibodies. In a preferred embodiment, such substances are employed for the treatment or prevention of certain types of neoplastic disorders associated such as cancer or hematological malignancies such as leukemia. In such embodiments, where an increased level of expression or activity of the present proteins is correlated with the presence of a disease such as cancer, the disease can be treated or prevented using any agent that can provoke a decrease in the level of activity or expression of the protein, such as antibodies, antisense molecules, ribozymes, dominant negative forms of the protein, compounds that inhibit the expression or activity of the proteins, and others. Alternatively, in cases where a decreased level of expression or activity of the proteins is correlated with the presence of a disease such as cancer, the disease can be treated using any agent that can cause an increase in the expression or activity of the protein, such as polynucleotides encoding the proteins, purified forms of the proteins, or any compound that causes an increase in the expression or activity of the proteins. Further, any detection of a correlation between the level of expression or activity of the protein and the presence or absence of a disease can be used to develop diagnostic or screening tools for the detection of the disease itself, or of a predisposition for the disease.

Uses of antibodies

Antibodies of the present invention have uses that include, but are not limited to, methods known in the art to purify, detect, and target the polypeptides of the present invention including both *in vitro* and *in vivo* diagnostic and therapeutic methods. An example of such use using immunoaffinity chromatography is given below. The antibodies of the present invention may be used either alone or in combination with other compositions. For example, the antibodies have use in immunoassays for qualitatively and quantitatively measuring levels of antigen-bearing substances, including the polypeptides of the present invention, in biological samples (*See, e.g.*, Harlow *et al.*, 1988). (Incorporated by reference in the entirety). The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

The invention further relates to antibodies that act as agonists or antagonists of the polypeptides of the present invention. For example, the present invention includes antibodies that

disrupt the receptor/ligand interactions with the polypeptides of the invention either partially or fully. Included are both receptor-specific antibodies and ligand-specific antibodies. Included are receptor-specific antibodies, which do not prevent ligand binding but prevent receptor activation. Receptor activation (i.e., signaling) may be determined by techniques described herein or otherwise known in the art. Also include are receptor-specific antibodies which both prevent ligand binding and receptor activation. Likewise, included are neutralizing antibodies that bind the ligand and prevent binding of the ligand to the receptor, as well as antibodies that bind the ligand, thereby preventing receptor activation, but do not prevent the ligand from binding the receptor. Further included are antibodies that activate the receptor. These antibodies may act as agonists for either all or less than all of the biological activities affected by ligand-mediated receptor activation. The antibodies may be specified as agonists or antagonists for biological activities comprising specific activities disclosed herein. The above antibody agonists can be made using methods known in the art. *See e.g.*, WO 96/40281; US Patent 5,811,097; Deng *et al.* (1998); Chen *et al.* (1998); Harrop *et al.* (1998); Zhu *et al.* (1998); Yoon *et al.* (1998); Prat *et al.* (1998); Pitard *et al.* (1997); Liautard *et al.* (1997); Carlson *et al.* (1997); Taryman *et al.* (1995); Muller *et al.* (1998); Bartunek *et al.* (1996) (said references incorporated by reference in their entireties).

As discussed above, antibodies of the polypeptides of the invention can, in turn, be utilized to generate anti-idiotypic antibodies that "mimic" polypeptides of the invention using techniques well known to those skilled in the art (*See, e.g.* Greenspan and Bona (1989) and Nissinoff (1991), which disclosures are hereby incorporated by reference in their entireties). For example, antibodies which bind to and competitively inhibit polypeptide multimerization or binding of a polypeptide of the invention to ligand can be used to generate anti-idiotypes that "mimic" the polypeptide multimerization or binding domain and, as a consequence, bind to and neutralize polypeptide or its ligand. Such neutralization anti-idiotypic antibodies can be used to bind a polypeptide of the invention or to bind its ligands/receptors, and thereby block its biological activity.

Immunoaffinity Chromatography

Antibodies prepared as described herein are coupled to a support. Preferably, the antibodies are monoclonal antibodies, but polyclonal antibodies may also be used. The support may be any of those typically employed in immunoaffinity chromatography, including Sepharose CL-4B (Pharmacia, Piscataway, NJ), Sepharose CL-2B (Pharmacia, Piscataway, NJ), Affi-gel 10 (Biorad, Richmond, CA), or glass beads.

The antibodies may be coupled to the support using any of the coupling reagents typically used in immunoaffinity chromatography, including cyanogen bromide. After coupling the antibody to the support, the support is contacted with a sample which contains a target polypeptide whose isolation, purification or enrichment is desired. The target polypeptide may be a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides

included in SEQ ID Nos: 242-272 and 274-384 as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool, variants and fragments thereof, or a fusion protein comprising said selected polypeptide or a fragment thereof.

Preferably, the sample is placed in contact with the support for a sufficient amount of time and under appropriate conditions to allow at least 50% of the target polypeptide to specifically bind to the antibody coupled to the support.

Thereafter, the support is washed with an appropriate wash solution to remove polypeptides which have non-specifically adhered to the support. The wash solution may be any of those typically employed in immunoaffinity chromatography, including PBS, Tris-lithium chloride buffer (0.1M lysine base and 0.5M lithium chloride, pH 8.0), Tris-hydrochloride buffer (0.05M Tris-hydrochloride, pH 8.0), or Tris/Triton/NaCl buffer (50mM Tris.cl, pH 8.0 or 9.0, 0.1% Triton X-100, and 0.5MNaCl).

After washing, the specifically bound target polypeptide is eluted from the support using the high pH or low pH elution solutions typically employed in immunoaffinity chromatography. In particular, the elution solutions may contain an eluant such as triethanolamine, diethylamine, calcium chloride, sodium thiocyanate, potasssium bromide, acetic acid, or glycine. In some embodiments, the elution solution may also contain a detergent such as Triton X-100 or octyl-beta-D-glucoside.

Import vectors

The GENSET polypeptides of the invention may also be used as a carrier to import a protein or peptide of interest, so-called cargo, into tissue-culture cells or in host organisms. A hydrophobic region of a GENSET polypeptide or a fragment thereof, preferably the signal peptide of a sequence selected from the group consisting of of SEQ ID Nos: 1-31 and 33-143 and clones inserts of the deposited clone pool, more preferably the short core hydrophobic region (h) of signal peptides may be used as a carrier.

When cell permeable peptides of limited size (approximately up to 25 amino acids) are to be translocated across cell membrane, chemical synthesis may be used in order to add the h region to either the C-terminus or the N-terminus to the cargo peptide of interest. Alternatively, when longer peptides or proteins are to be imported into cells, nucleic acids can be genetically engineered, using techniques familiar to those skilled in the art, in order to link the cDNA sequence or fragment thereof encoding the hydrophobic region to the 5' or the 3' end of a DNA sequence coding for a cargo polypeptide. Such genetically engineered nucleic acids are then translated either *in vitro* or *in vivo* after transfection into appropriate cells, using conventional techniques to produce the resulting cell permeable polypeptide. Suitable hosts cells are then simply incubated with the cell permeable polypeptide which is then translocated across the membrane.

This method may be applied to study diverse intracellular functions and cellular processes. For instance, it has been used to probe functionally relevant domains of intracellular proteins and to examine protein-protein interactions involved in signal transduction pathways (Lin *et al., J. Biol. Chem.*, 270: 14225-14258 (1995); Lin *et al., J. Biol. Chem.*, 271: 5305-5308 (1996); Rojas *et al., J. Biol. Chem.*, 271: 27456-27461 (1996); Rojas *et al., Nature Biotech.*, 16: 370-375 (1998); Liu *et al., Proc. Natl. Acad. Sci. USA*, 93: 11819-11824 (1996); Rojas *et al., Bioch. Biophys. Res. Commun.*, 234: 675-680 (1997) Du *et al., J. Peptide Res.*, 51: 235-243 (1998)).

Such techniques may be used in cellular therapy to import proteins producing therapeutic effects. For instance, cells isolated from a patient may be treated with imported therapeutic proteins and then re-introduced into the host organism.

Alternatively, the hydrophobic region of signal peptides of the present invention could be used in combination with a nuclear localization signal to deliver nucleic acids into cell nucleus. Such oligonucleotides may be antisense oligonucleotides or oligonucleotides designed to form triple helixes, as described herein, in order to respectively inhibit processing or maturation of a target cellular RNA.

EXPRESSION OF GENSET PRODUCTS

Spatial expression of the GENSET genes of the invention

Tissue expression of the cDNAs of the present invention was examined. Table IX list the Genset's libraries of tissues and cell types examined that express the polynucleotides of the present invention. The tissues and cell types examined for polynucleotide expression were: adrenal gland (AG), bone marrow (BM), brain (Br), cancerous protate (CP), cerebellum (Ce), colon (Co), dystrophic muscle (DM), fetal brain (FB), fetal kidney (FK), fetal liver (FL), heart (He), hypertrophic prostate (HP), kidney (Ki), liver (Li), lung (Lu), lung cells (LC), lymph ganglia (LG), lymphocytes (Ly), muscle (Mu), Ovary (Ov), pancreas (Pa), pituitary gland (PG), placenta (Pl), prostate (Pr), salivary gland (SG), spinal cord (SC), spleen (Sp), stomach/intestine (SI), substantia nigra (SN), testis (Te), thyroid (Ty), umbilical cord (UC) and uterus (Ut).

For each cDNA referred to by its sequence identification number (first column), the number of proprietary 5'ESTs (i.e. cDNA fragments) expressed in a particular tissue referred to by its name is indicated after a semi column (second column). In addition, the bias in the spatial distribution of the polynucleotide sequences of the present invention was examined by comparing the relative proportions of the biological polynucleotides of a given tissue using the following statistical analysis. The under- or over-representation of a polynucleotide of a given cluster in a given tissue was performed using the normal approximation of the binomial distribution. When the observed proportion of a polynucleotide of a given tissue in a given consensus had less than 1% chance to occur randomly according to the chi2 test, the frequency bias was reported as "preferred". The results are given in Table X as follows. For each polynucleotide showing a bias in tissue distribution

as referred to by its sequence identification number in the first column, the list of tissues where the polynucleotides are under-represented is given in the second column entitled "low frequency expression" and the list of tissues where the polynucleotides are over-represented is given in the third column entitled "high frequency expression".

The cellular localization of some polypeptides of the invention was also determined using the "psort software" (Nakai, and Horton, (1999); Nakai and Kanehisa, (1992), which disclosures are hereby incorporated by reference in their entireties). For each polypeptide identified by its sequence identification number in the first column, the second column of Table XI list the predicted subcellular localization.

Evaluation of Expression Levels and Patterns of GENSET mRNAs

The spatial and temporal expression patterns of GENSET mRNAs, as well as their expression levels, may also be further determined as follows.

Expression levels and patterns of GENSET mRNAs may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277, the entire contents of which are hereby incorporated by reference. Briefly, a GENSET polynucleotide, or fragment thereof corresponding to the gene encoding the mRNA to be characterized is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the GENSET polynucleotide is at least a 100 nucleotides in length. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (i.e. biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (i.e. RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

The GENSET cDNAs, or fragments thereof may also be tagged with nucleotide sequences for the serial analysis of gene expression (SAGE) as disclosed in UK Patent Application No. 2 305 241 A, the entire contents of which are incorporated by reference. In this method, cDNAs are prepared from a cell, tissue, organism or other source of nucleic acid for which it is desired to determine gene expression patterns. The resulting cDNAs are separated into two pools. The cDNAs in each pool are cleaved with a first restriction endonuclease, called an "anchoring enzyme," having a recognition site which is likely to be present at least once in most cDNAs. The fragments which contain the 5' or 3' most region of the cleaved cDNA are isolated by binding to a capture

medium such as streptavidin coated beads. A first oligonucleotide linker having a first sequence for hybridization of an amplification primer and an internal restriction site for a "tagging endonuclease" is ligated to the digested cDNAs in the first pool. Digestion with the second endonuclease produces short "tag" fragments from the cDNAs. A second oligonucleotide having a second sequence for hybridization of an amplification primer and an internal restriction site is ligated to the digested cDNAs in the second pool. The cDNA fragments in the second pool are also digested with the "tagging endonuclease" to generate short "tag" fragments derived from the cDNAs in the second pool. The "tags" resulting from digestion of the first and second pools with the anchoring enzyme and the tagging endonuclease are ligated to one another to produce "ditags." In some embodiments, the ditags are concatamerized to produce ligation products containing from 2 to 200 ditags. The tag sequences are then determined and compared to the sequences of the GENSET cDNAs to determine which genes are expressed in the cell, tissue, organism, or other source of nucleic acids from which the tags were derived. In this way, the expression pattern of a GENSET gene in the cell, tissue, organism, or other source of nucleic acids is obtained.

Quantitative analysis of GENSET gene expression may also be performed using arrays. For example, quantitative analysis of gene expression may be performed with GENSET polynucleotides, or fragments thereof in a complementary DNA microarray as described by Schena *et al.* (1995 and 1996) which disclosures are hereby incorporated by reference in their entireties. GENSET cDNAs or fragments thereof are amplified by PCR and arrayed from 96-well microtiter plates onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C. Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm^2 microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1X SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1X SSC/0.2% SDS). Arrays are scanned in 0.1X SSC using a fluorescence laser scanning device fitted with a custom filter set. Accurate differential expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of the expression of genes may also be performed with GENSET cDNAs or fragments thereof in complementary DNA arrays as described by Pietu *et al.* (1996), which disclosure is hereby incorporated by reference in its entirety. The GENSET polynucleotides of the invention or fragments thereof are PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by

phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis of GENSET genes can be done through high density nucleotide arrays as described by Lockhart *et al.* (1996) and Sosnowski *et al.* (1997), which disclosures are hereby incorporated by reference in their entireties. Oligonucleotides of 15-50 nucleotides corresponding to sequences of a GENSET polynucleotide or fragments thereof are synthesized directly on the chip (Lockhart *et al.*, supra) or synthesized and then addressed to the chip (Sosnowski *et al.*, supra). Preferably, the oligonucleotides are about 20 nucleotides in length. cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart *et al.*, (supra) and application of different electric fields (Sosnowsky *et al.*, supra), the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of the GENSET mRNA.

Uses of GENSET expression data

Once the expression levels and patterns of a GENSET mRNA has been determined using any technique known to those skilled in the art, in particular those described in the section entitled "Evaluation of Expression Levels and Patterns of GENSET mRNAs", or using the instant disclosure, these information may be used to design GENSET specific markers for detection, identification, screening and diagnosis purposes as well as to design DNA constructs with an expression pattern similar to a GENSET expression pattern.

Detection of GENSET expression and/or biological activity

The invention further relates to methods of detection of GENSET expression and/or biological activity in a biological sample using the polynucleotide and polypeptide sequences described herein. Such method scan be used, for example, as a screen for normal or abnormal GENSET expression and/or biological activity and, thus, can be used diagnostically. The biological sample for use in the methods of the present invention includes a suitable sample from, for example, a mammal, particularly a human. For example, the sample can be issued from tissues or cell lines having the same origin as tissues or cell lines in which the polypeptide is known to be expressed using the data from Table IX.

Detection of GENSET products

The invention further relates to methods of detection of GENSET polynucleotides or polypeptides in a sample using the sequences described herein and any techniques known to those

skilled in the art. For example, a labeled polynucleotide probe having all or a functional portion of the nucleotide sequence of a GENSET polynucleotide can be used in a method to detect a GENSET polynucleotide in a sample. In one embodiment, the sample is treated to render the polynucleotides in the sample available for hybridization to a polynucleotide probe, which can be DNA or RNA. The resulting treated sample is combined with a labeled polynucleotide probe having all or a portion of the nucleotide sequence of the GENSET cDNA or genomic sequence, under conditions appropriate for hybridization of complementary sequences to occur. Detection of hybridization of polynucleotides from the sample with the labeled nucleic probe indicates the presence of GENSET polynucleotides in a sample. The presence of GENSET mRNA is indicative of GENSET expression.

Consequently, the invention comprises methods for detecting the presence of a polynucleotide comprising a nucleotide sequence selected from a group consisting of the sequences of SEQ ID Nos: 1-241, the sequences of clone inserts of the deposited clone pool, sequences fully complementary thereto, fragments and variants thereof in a sample. In a first embodiment, said method comprises the following steps of:

a) bringing into contact said sample and a nucleic acid probe or a plurality of nucleic acid probes which hybridize to said selected nucleotide sequence; and

b) detecting the hybrid complex formed between said probe or said plurality of probes and said polynucleotide.

In a preferred embodiment of the above detection method, said nucleic acid probe or said plurality of nucleic acid probes is labeled with a detectable molecule. In another preferred embodiment of the above detection method, said nucleic acid probe or said plurality of nucleic acid probes has been immobilized on a substrate. In still another preferred embodiment, said nucleic acid probe or said plurality of nucleic acid probes has a sequence comprised in a sequence complementary to said selected sequence.

In a second embodiment, said method comprises the following steps of:

a) contacting said sample with amplification reaction reagents comprising a pair of amplification primers located on either side of the region of said nucleotide sequence to be amplified;

b) performing an amplification reaction to synthesize amplification products containing said region of said selected nucleotide sequence; and

c) detecting said amplification products.

In a preferred embodiment of the above detection method, when the polynucleotide to be amplified is a RNA molecule, preliminary reverse transcription and synthesis of a second cDNA strand are necessary to provide a DNA template to be amplified. In another preferred embodiment of the above detection method, the amplification product is detected by hybridization with a labeled probe having a sequence which is complementary to the amplified region. In still another preferred

embodiment, at least one of said amplification primer has a sequence comprised in said selected sequence or in the sequence complementary to said selected sequence.

Alternatively, a method of detecting GENSET expression in a test sample can be accomplished using any product which binds to a GENSET polypeptide of the present invention or a portion of a GENSET polypeptide. Such products may be antibodies, binding fragments of antibodies, polypeptides able to bind specifically to GENSET polypeptides or fragments thereof, including GENSET agonists and antagonists. Detection of specific binding to the antibody indicates the presence of a GENSET polypeptide in the sample (e.g., ELISA).

Consequently, the invention is also directed to a method for detecting specifically the presence of a GENSET polypeptide according to the invention in a biological sample, said method comprising the following steps of:

a) bringing into contact said biological sample with a product able to bind to a polypeptide of the invention or fragments thereof;

b) allowing said product to bind to said polypeptide to form a complex; and

b) detecting said complex.

In a preferred embodiment of the above detection method, the product is an antibody. In a more preferred embodiment, said antibody is labeled with a detectable molecule. In another more preferred embodiment of the above detection method, said antibody has been immobilized on a substrate.

In addition, the invention also relates to methods of determining whether a GENSET product (e.g. a polynucleotide or polypeptide) is present or absent in a biological sample, said methods comprising the steps of:

a) obtaining said biological sample from a human or non-human animal, preferably a mammal;

b) contacting said biological sample with a product able to bind to a GENSET polynucleotide or polypeptide of the invention; and

c) determining the presence or absence of said GENSET product in said biological sample.

Compounds that specifically binds a GENSET product may either be compounds binding to a GENSET polypeptide (e.g. binding proteins, antibodies or binding fragments thereof (e.g. F(ab')2 fragments) or compounds bindint to a GENSET polynucleotide (e.g. a complementary probe or primer).

The present invention also relates to kits that can be used in the detection of GENSET expression products. The kit can comprise a compound that specifically binds a GENSET polypeptide (e.g. binding proteins, antibodies or binding fragments thereof (e.g. F(ab')2 fragments) or a GENSET mRNA (e.g. a complementary probe or primer), for example, disposed within a container means. The kit can further comprise ancillary reagents, including buffers and the like.

Detection of a GENSET biological activity

The invention further includes methods of detecting specifically a GENSET biological activity. Assessing the GENSET biological activity may be performed using a variety of techniques, including those described in the section entitled "**Erreur! Source du renvoi introuvable.**".

Consequently, the invention is directed to a method for detecting specifically GENSET biological activity in a biological sample, said method comprising the following steps:

a) obtaining a biological sample from a human or non-human mammal; and

b) detecting a GENSET biological activity.

The present invention also relates to kits that can be used in the detection of GENSET biological activity.

Identification of a specific context of GENSET expression

When the expression pattern of a GENSET mRNA shows that a GENSET gene is specifically expressed in a given context, probes and primers specific for this gene as well as antibodies binding to the GENSET polynucleotide may then be used as markers for a specific context. Examples of specific contexts are: specific expression in a given tissue/cell or tissue/cell type, expression at a given stage of development of a process such as embryo development or disease development, or specific expression in a given organelle. Such primers, probes, and antibodies are useful commercially to identify tissues/cells/organelles of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using any technique known to those skilled in the art including *in situ* PCR or immunochemistry for example.

For example, the cDNAs and proteins of the sequence listing and fragments thereof, may be used to distinguish human tissues/cells from non-human tissues/cells and to distinguish between human tissues/cells/organelles that do and do not express the polynucleotides comprising the cDNAs. By knowing the expression pattern of a given GENSET, either through routine experimentation or by using the instant disclosure, the polynucleotides and polypeptides of the present invention may be used in methods of determining the identity of an unknown tissue/cell sample/organelle. As part of determining the identity of an unknown tissue/cell sample/organelle, the polynucleotides and polypeptides of the present invention may be used to determine what the unknown tissue/cell sample is and what the unknown sample is not. For example, if a cDNA is expressed in a particular tissue/cell type/organelle, and the unknown tissue/cell sample/organelle does not express the cDNA, it may be inferred that the unknown tissue/cells are either not human or not the same human tissue/cell type/organelle as that which expresses the cDNA. These methods of determining tissue/cell/organelle identity are based on methods which detect the presence or absence of the mRNA (or corresponding cDNA) in a tissue/cell sample using methods well know in the art (e.g., hybridization, PCR based methods, immunoassays, immunochemistry, ELISA).

Examples of such techniques are described in more detail below. Therefore, the invention encompasses uses of the polynucleotides and polypeptides of the invention as tissue markers. In a preferred embodiment, polynucleotides preferentially expressed in given tissues as indicated in Table X and polypeptides encoded by such polynucleotides are used for this purpose. The invention also encompasses uses of polypeptides of the invention as organelle markers. In a preferred embodiment, polypeptides preferentially expressed in given subcellular compartment as indicated in Table XI are used for this purpose.

Consequently, the present invention encompasses methods of identification of a tissue/cell type/subcellular compartment, wherein said method includes the steps of:

a) contacting a biological sample which identity is to be assayed with a product able to bind a GENSET product; and

b) determining whether a GENSET product is expressed in said biological sample.

Products that are able to bind specifically to a GENSET product, namely a GENSET polypeptide or a GENSET mRNA, include GENSET binding proteins, antibodies or binding fragments thereof (e.g. F(ab')2 fragments), as well as GENSET complementary probes and primers.

Step b) may be performed using any detection method known to those skilled in the art including those disclosed herein, especially in the section entitled "Detection of GENSET expression and/or biological activity"..

Identification of Tissue Types or Cell Species by Means of Labeled Tissue Specific Antibodies

Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of antibody preparations which are conjugated, directly (e.g., green fluorescent protein) or indirectly to a detectable marker. Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative interpretation.

Antisera for these procedures must have a potency exceeding that of the native preparation, and for that reason, antibodies are concentrated to a mg/ml level by isolation of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate fractionation. Also, to provide the most specific antisera, unwanted antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble immunoabsorbents, before the antibodies are labeled with the marker. Either monoclonal or heterologous antisera is suitable for either procedure.

A. Immunohistochemical Techniques

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, (1980) or Rose *et al.*, (1980), which disclosures are hereby incorporated by reference in their entireties.

A fluorescent marker, either fluorescein or rhodamine, is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such as horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific anti-tissue antibodies can be labeled with ferritin or other electron dense particles, and localization of the ferritin coupled antigen-antibody complexes achieved by means of an electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example ^{125}I, and detected by overlaying the antibody treated preparation with photographic emulsion. Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single protein or peptide identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently or in mixtures, as required. Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4 um, unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations to provide a positive control, a negative control, for example, pre-immune sera, and a control for non-specific staining, for example, buffer. Treated sections are incubated in a humid chamber for 30 min at room temperature, rinsed, then washed in buffer for 30-45 min. Excess fluid is blotted away, and the marker developed. If the tissue specific antibody was not labeled in the first incubation, it can be labeled at this time in a second antibody-antibody reaction, for example, by adding fluorescein- or enzyme-conjugated antibody against the immunoglobulin class of the antiserum-producing species, for example, fluorescein labeled antibody to mouse IgG. Such labeled sera are commercially available. The antigen found in the tissues by the above procedure can be quantified by measuring the intensity of color or fluorescence on the tissue section, and calibrating that signal using appropriate standards.

B. Identification of Tissue Specific Soluble Proteins

The visualization of tissue specific proteins and identification of unknown tissues from that procedure is carried out using the labeled antibody reagents and detection strategy as described for immunohistochemistry; however the sample is prepared according to an electrophoretic technique to distribute the proteins extracted from the tissue in an orderly array on the basis of molecular weight for detection. A tissue sample is homogenized using a Virtis apparatus; cell suspensions are disrupted by Dounce homogenization or osmotic lysis, using detergents in either case as required to disrupt cell membranes, as is the practice in the art. Insoluble cell components such as nuclei, microsomes, and membrane fragments are removed by ultracentrifugation, and the soluble protein-containing fraction concentrated if necessary and reserved for analysis. A sample of the soluble protein solution is resolved into individual protein species by conventional SDS polyacrylamide

electrophoresis as described, for example, by Davis *et al.*, Section 19-2 (1986), using a range of amounts of polyacrylamide in a set of gels to resolve the entire molecular weight range of proteins to be detected in the sample. A size marker is run in parallel for purposes of estimating molecular weights of the constituent proteins. Sample size for analysis is a convenient volume of from 5 to55 ul, and containing from about 1 to 100 ug protein. An aliquot of each of the resolved proteins is transferred by blotting to a nitrocellulose filter paper, a process that maintains the pattern of resolution. Multiple copies are prepared. The procedure, known as Western Blot Analysis, is well described in Davis *et al.*, (1986) Section 19-3. One set of nitrocellulose blots is stained with Coomassie Blue dye to visualize the entire set of proteins for comparison with the antibody bound proteins. The remaining nitrocellulose filters are then incubated with a solution of one or more specific antisera to tissue specific proteins prepared as described herein. In this procedure, as in procedure A above, appropriate positive and negative sample and reagent controls are run.

In either procedure A or B, a detectable label can be attached to the primary tissue antigen-primary antibody complex according to various strategies and permutations thereof. In a straightforward approach, the primary specific antibody can be labeled; alternatively, the unlabeled complex can be bound by a labeled secondary anti-IgG antibody. In other approaches, either the primary or secondary antibody is conjugated to a biotin molecule, which can, in a subsequent step, bind an avidin conjugated marker. According to yet another strategy, enzyme labeled or radioactive protein A, which has the property of binding to any IgG, is bound in a final step to either the primary or secondary antibody. The visualization of tissue specific antigen binding at levels above those seen in control tissues to one or more tissue specific antibodies, prepared from the gene sequences identified from cDNA sequences, can identify tissues of unknown origin, for example, forensic samples, or differentiated tumor tissue that has metastasized to foreign bodily sites.

Targeting of compounds to subcellular compartments

GENSET Polypeptides expressed in specific cellular compartments/organelels may also be used to target compounds to these compartments/organelles. The invention therefore encompasses uses of polypeptides and polynucleotides of the invention as organelle targeting tools.

In a first embodiment, GENSET polypeptides expressed in mitochondria may be used to target heterologous compounds, either polypeptides or polynucleotides to mitochondria by recombinantly or chemically fusing a fragment of the protein of the invention to an heterologous polypeptide or polynucleotide. Preferred fragments are signal peptide, amphiphilic alpha helices and/or any other fragments of the protein of the invention, or part thereof, that may contain targeting signals for mitochondria including but not limited to matrix targeting signals as defined in Herrman and Neupert, (2000); Bhagwat *et al.* (1999), Murphy (1997); Glaser *et al.* (1998); Ciminale *et al.* (1999), which disclosures are hereby incorporated by reference in their entireties. Such heterologous compounds may be used to modulate mitochondria's activities. For example,

they may be used to induce and/or prevent mitochondrial-induced apoptosis or necrosis. In addition, heterologous polynucleotides may be used for mitochondrial gene therapy to replace a defective mitochondrial gene and/or to inhibit the deleterious expression of a mitochondrial gene.

In a second embodiment, GENSET polypeptides expressed in the endoplasmic reticulum may be used to target heterologous polypeptides to the endoplasmic reticulum by recombinantly or chemically fusing a fragment of the proteins of the invention to an heterologous polypeptide. Preferred fragments are any fragments of the proteins of the invention, or part thereof, that may contain targeting signals for the endoplasmic reticulum such as those described in Pidoux and Armstrong (1992), Munro and Pelham (1987); Pelham (1990), which disclosures are hereby incorporated by reference in their entireties.

In a third embodiment, GENSET polypeptides expressed in the nucleus may be used to target heterologous polypeptides or polynucleotides to the nucleus by recombinantly or chemically fusing a fragment of the proteins or polynuleotide of the invention to an heterologous polypeptide or polynucleotide. Preferred fragments are any fragments of the proteins or polynuclotide of the invention, or part thereof, that may contain targeting signals for the nucleus (nuclear localization signals) such as those described in Christophe et al. (2000), which disclosure is hereby incorporated by reference in its entirety.

In a fourth embodiment, GENSET polypeptides expressed in the nucleus may be used to target heterologous polypeptides to the Golgi apparatus by recombinantly or chemically fusing a fragment of the protein of the invention to an heterologous polypeptide. Preferred fragments are signal peptide, transmembrane domains, tyrosine containing regions and/or any other fragments of the proteins of the invention, or part thereof, that may contain (1) targeting signals for the Golgi apparatus such as the ones described in Ugur and Jones, (2000); Picetti and Borrelli, (2000), (2) tyrosine-based Golgi targeting signal region (Zhan et al., (1998); Watson and Pessin (2000); Ward and Moss (2000), or (3) any other region as defined in Munro, (1998); Luetterforst et al., (1999); Essl et al., (1999), which disclosures are hereby incorporated by reference in their entireties.

Screening and diagnosis of abnormal GENSET expression and/or biological activity

Moreover, antibodies and/or primers specific for GENSET expression may also be used to identify abnormal GENSET expression and/or biological activity, and subsequently to screen and/or diagnose disorders associated with abnormal GENSET expression. For example, a particular disease may result from lack of expression, over expression, or under expression of a GENSET mRNA. By comparing mRNA expression patterns and quantities in samples taken from healthy individuals with those from individuals suffering from a particular disorder, genes responsible for this disorder may be identified. Primers, probes and antibodies specific for this GENSET may then be used to elaborate kits of screening and diagnosis for a disorder in which the gene of interest is

specifically expressed or in which its expression is specifically dysregulated, i.e. underexpressed or overexpressed.

Screening for specific disorders

The present invention also relates to methods of identifying individuals having elevated or reduced levels of GENSET, which individuals are likely to benefit from therapies to suppress or enhance GENSET expression, respectively. One example of such methods comprises the steps of:

a) obtaining from a human or non-human mammal a biological sample;

b) detecting the presence in said sample of a GENSET product (mRNA or protein) using any method known to those skilled in the art including those described herein, especially at the section entitled "Detection of GENSET products";

c) comparing the amount of said GENSET product present in said sample with that of a control sample; and

d) determing whether said human or non-human mammal has a reduced or elevated level of GENSET expression compared to the control sample.

Such individuals with reduced or elevated levels of GENSET products may be predisposed to disorders associated with dyregulation of GENSET gene expression and thus would be candidates for therapies. The identification of elevated levels of GENSET in a patient would be indicative of an individual that would benefit from treatment with agents that suppress GENSET expression or activity. The identification of low levels of GENSET in a patient would be indicative of an individual that would benefit from agents that induce GENSET expression or activity.

Biological samples suitable for use in this method include biological fluids such as blood, lymph, saliva, sperm, maternal milk, and tissue samples (e.g. biopsies) as well as cell cultures or cell extracts derived, for example, from tissue biopsies. The detection step of the present method can be performed using standard protocols for protein/mRNA detection. Examples of suitable protocols include Northern blot analysis, immunoassays (e.g. RIA, Western blots, immunohistochemical analyses), and PCR.

Thus, the present invention further relates to methods of identifying individuals or non-human animals at increased risk for developing, or present state of having, certain diseases/disorders associated with GENSET abnormal expression or biological activity. One example of such methods comprises the steps of:

a) obtaining from a human or non-human mammal a biological sample;

b) detecting the presence or absence in said sample of a GENSET product (mRNA or protein);

c) comparing the amount of said GENSET product present in said sample with that of a control sample; and

d) determing whether said human or non-human mammal is at increased risk for developing, or present state of having, a diseases or disorder.

In accordance with this method, the presence in the sample of altered levels of GENSET product indicates that the subject is predisposed to the above-indicated diseases/disorders. Biological samples suitable for use in this method include biological fluids such as blood, lymph, saliva, sperm, maternal milk, and tissue samples (e.g. biopsies.

The diagnostic methodologies described herein are applicable to both humans and non-human mammals.

Detection of GENSET mutations

The invention also encompasses methods to detect mutations in GENSET polynucleotides of the invention. Such methods may advantageously be used to detect mutations occurring in GENSET genes and preferably in their regulatory regions. When the mutation was proven to be associated with a disease, screening for such mutations may be used for screening and diagnosis purposes.

In one embodiment of the oligonucleotide arrays of the invention, an oligonucleotide probe matrix may advantageously be used to detect mutations occurring in GENSET genes and preferably in their regulatory regions. For this particular purpose, probes are specifically designed to have a nucleotide sequence allowing their hybridization to the genes that carry known mutations (either by deletion, insertion or substitution of one or several nucleotides). By known mutations, it is meant, mutations on the GENSET genes that have been identified according, for example to the technique used by Huang *et al.*(1996) or Samson *et al.*(1996), which disclosures are hereby incorporated by reference in their entireties.

Another technique that is used to detect mutations in GENSET genes is the use of a high-density DNA array. Each oligonucleotide probe constituting a unit element of the high density DNA array is designed to match a specific subsequence of a GENSET genomic DNA or cDNA. Thus, an array consisting of oligonucleotides complementary to subsequences of the target gene sequence is used to determine the identity of the target sequence with the wild gene sequence, measure its amount, and detect differences between the target sequence and the reference wild gene sequence of the GENSET gene. In one such design, termed 4L tiled array, is implemented a set of four probes (A, C, G, T), preferably 15-nucleotide oligomers. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. Consequently, a nucleic acid target of length L is scanned for mutations with a tiled array containing 4L probes, the whole probe set containing all the possible mutations in the known wild reference sequence. The hybridization signals of the 15-mer probe set tiled array are perturbed by a single base change in the target sequence. As a consequence, there is a characteristic loss of signal or a "footprint" for the probes

flanking a mutation position. This technique was described by Chee *et al.* in 1996, which disclosure is hereby incorporated by reference in its entirety.

Construction of DNA constructs with a GENSET expression pattern

In addition, characterization of the spatial and temporal expression patterns and expression levels of GENSET mRNAs is also useful for constructing expression vectors capable of producing a desired level of gene product in a desired spatial or temporal manner, as discussed below.

DNA Construct That Enables Directing Temporal And Spatial GENSET Gene Expression In Recombinant Cell Hosts And In Transgenic Animals.

In order to study the physiological and phenotypic consequences of a lack of synthesis of a GENSET protein, both at the cell level and at the multi cellular organism level, the invention also encompasses DNA constructs and recombinant vectors enabling a conditional expression of a specific allele of a GENSET genomic sequence or cDNA and also of a copy of this genomic sequence or cDNA harboring substitutions, deletions, or additions of one or more bases as regards to a nucleotide sequence selected from the group consisting of sequences of SEQ ID Nos 1-241 and sequences of clone inserts of the deposited clone pool, or a fragment thereof, these base substitutions, deletions or additions being located either in an exon, an intron or a regulatory sequence, but preferably in the 5'-regulatory sequence or in an exon of the GENSET genomic sequence or within the GENSET cDNA.

A first preferred DNA construct is based on the tetracycline resistance operon *tet* from *E. coli* transposon Tn10 for controlling the GENSET gene expression, such as described by Gossen *et al.*(1992, 1995) and Furth *et al.*(1994), which disclosures are hereby incorporated by reference in their entireties. Such a DNA construct contains seven *tet* operator sequences from Tn10 (*tet*op) that are fused to either a minimal promoter or a 5'-regulatory sequence of the GENSET gene, said minimal promoter or said GENSET regulatory sequence being operably linked to a polynucleotide of interest that codes either for a sense or an antisense oligonucleotide or for a polypeptide, including a GENSET polypeptide or a peptide fragment thereof. This DNA construct is functional as a conditional expression system for the nucleotide sequence of interest when the same cell also comprises a nucleotide sequence coding for either the wild type (tTA) or the mutant (rTA) repressor fused o the activating domain of viral protein VP16 of herpes simplex virus, placed under the control of a promoter, such as the HCMVIE1 enhancer/promoter or the MMTV-LTR. Indeed, a preferred DNA construct of the invention comprise both the polynucleotide containing the *tet* operator sequences and the polynucleotide containing a sequence coding for the tTA or the rTA repressor. In a specific embodiment, the conditional expression DNA construct contains the sequence encoding the mutant tetracycline repressor rTA, the expression of the polynucleotide of interest is silent in the absence of tetracycline and induced in its presence.

DNA Constructs Allowing Homologous Recombination: Replacement Vectors

A second preferred DNA construct will comprise, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the GENSET genomic sequence; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycine resistance (*neo*); and (c) a second nucleotide sequence that is comprised in the GENSET genomic sequence, and is located on the genome downstream the first GENSET nucleotide sequence (a).

In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream the nucleotide sequence (a) or downstream the nucleotide sequence (c). Preferably, the negative selection marker comprises the thymidine kinase (*tk*) gene (Thomas *et al.*, 1986), the hygromycine beta gene (Te Riele *et al.*, 1990), the *hprt* gene (Van der Lugt *et al.*, 1991; Reid *et al.*, 1990) or the Diphteria toxin A fragment (*Dt-A*) gene (Nada *et al.*, 1993; Yagi *et al.*1990), which disclosures are hereby incorporated by reference in their entireties. Preferably, the positive selection marker is located within a GENSET exon sequence so as to interrupt the sequence encoding a GENSET protein. These replacement vectors are described, for example, by Thomas *et al.*(1986; 1987), Mansour *et al.*(1988) and Koller *et al.*(1992).

The first and second nucleotide sequences (a) and (c) may be indifferently located within a GENSET regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The size of the nucleotide sequences (a) and (c) ranges from 1 to 50 kb, preferably from 1 to 10 kb, more preferably from 2 to 6 kb and most preferably from 2 to 4 kb.

DNA Constructs Allowing Homologous Recombination: Cre-LoxP System.

These new DNA constructs make use of the site specific recombination system of the P1 phage. The P1 phage possesses a recombinase called Cre which interacts specifically with a 34 base pairs *lox*P site. The *lox*P site is composed of two palindromic sequences of 13 bp separated by a 8 bp conserved sequence (Hoess *et al.*, 1986), which disclosure is hereby incorporated by reference in its entirety. The recombination by the Cre enzyme between two *lox*P sites having an identical orientation leads to the deletion of the DNA fragment.

The Cre-*lox*P system used in combination with a homologous recombination technique has been first described by Gu *et al.*(1993, 1994), which disclosures are hereby incorporated by reference in their entireties. Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two *lox*P sites in the same orientation and located at the respective ends of a nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the nucleus of the recombinant cell host. The recombinase enzyme may be brought at the desired time either by (a) incubating the recombinant cell hosts in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki *et al.*(1995), which disclosure is hereby

incorporated by reference in its entirety, or by lipofection of the enzyme into the cells, such as described by Baubonis *et al.*(1993), which disclosure is hereby incorporated by reference in its entirety; (b) transfecting the cell host with a vector comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter being optionally inducible, said vector being introduced in the recombinant cell host, such as described by Gu *et al.*(1993) and Sauer *et al.*(1988), which disclosures are hereby incorporated by reference in their entireties; (c) introducing in the genome of the cell host a polynucleotide comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host either by a random insertion event or an homologous recombination event, such as described by Gu *et al.*(1994).

In a specific embodiment, the vector containing the sequence to be inserted in the GENSET gene by homologous recombination is constructed in such a way that selectable markers are flanked by *lox*P sites of the same orientation, it is possible, by treatment by the Cre enzyme, to eliminate the selectable markers while leaving the GENSET sequences of interest that have been inserted by an homologous recombination event. Again, two selectable markers are needed: a positive selection marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the Cre-*lox*P system are described by Zou *et al.*(1994), which disclosure is hereby incorporated by reference in its entirety.

Thus, a third preferred DNA construct of the invention comprises, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the GENSET genomic sequence; (b) a nucleotide sequence comprising a polynucleotide encoding a positive selection marker, said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a *lox*P site, the two sites being placed in the same orientation; and (c) a second nucleotide sequence that is comprised in the GENSET genomic sequence, and is located on the genome downstream of the first GENSET nucleotide sequence (a).

The sequences defining a site recognized by a recombinase, such as a *lox*P site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for which the conditional excision is sought. In one specific embodiment, two *lox*P sites are located at each side of the positive selection marker sequence, in order to allow its excision at a desired time after the occurrence of the homologous recombination event.

In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two loxP sites, is performed at a desired time, due to the presence within the genome of the recombinant host cell of a sequence encoding the Cre enzyme operably linked to a promoter sequence, preferably an inducible promoter, more preferably a tissue-specific promoter sequence

and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu *et al.*(1994).

The presence of the Cre enzyme within the genome of the recombinant cell host may result from the breeding of two transgenic animals, the first transgenic animal bearing the GENSET-derived sequence of interest containing the *lox*P sites as described above and the second transgenic animal bearing the *Cre* coding sequence operably linked to a suitable promoter sequence, such as described by Gu *et al.*(1994).

Spatio-temporal control of the Cre enzyme expression may also be achieved with an adenovirus based vector that contains the Cre gene thus allowing infection of cells, or *in vivo* infection of organs, for delivery of the Cre enzyme, such as described by Anton and Graham (1995) and Kanegae *et al.*(1995), which disclosures are hereby incorporated by reference in their entireties.

The DNA constructs described above may be used to introduce a desired nucleotide sequence of the invention, preferably a GENSET genomic sequence or a GENSET cDNA sequence, and most preferably an altered copy of a GENSET genomic or cDNA sequence, within a predetermined location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination).

MODIFYING GENSET EXPRESSION AND/OR BIOLOGICAL ACTIVITY

Modifying endogenous GENSET expression and/or biological activity is expressly contemplated by the present invention.

Screening for compounds that modulate GENSET expression and/or biological activity

The present invention further relates to compounds able to modulate GENSET expression and/or biological activity and methods to use these compounds. Such compounds may interact with the regulatory sequences of GENSET genes or they may interact with GENSET polypeptides directly or indirectly.

Compounds Interacting With GENSET Regulatory Sequences

The present invention also concerns a method for screening substances or molecules that are able to interact with the regulatory sequences of a GENSET gene, such as for example promoter or enhancer sequences in untranscribed regions of the genomic DNA, as determined using any techniques known to those skilled in the art including those described in the section entitled "Identification of Promoters in Cloned Upstream Sequences, or such as regulatory sequences located in untranslated regions of GENSET mRNA.

Sequences within untranscribed or untranslated regions of polynucleotides of the invention may be identified by comparison to databases containing known regulatory sequence such as

transcription start sites, transcription factor binding sites, promoter sequences, enhancer sequences, 5'UTR and 3'UTR elements (Pesole *et al.*, 2000; http://igs-server.cnrs-mrs.fr/~gauthere/UTR/index.html). Alternatively, the regulatory sequences of interest may be identified through conventional mutagenesis or deletion analyses of reporter plasmids using, for instance, techniques described in the section entitled "Identification of Promoters in Cloned Upstream Sequences".

Following the identification of potential GENSET regulatory sequences, proteins which interact with these regulatory sequences may be identified as described below.

Gel retardation assays may be performed independently in order to screen candidate molecules that are able to interact with the regulatory sequences of the GENSET gene, such as described by Fried and Crothers (1981), Garner and Revzin (1981) and Dent and Latchman (1993), the teachings of these publications being herein incorporated by reference. These techniques are based on the principle according to which a DNA or mRNA fragment which is bound to a protein migrates slower than the same unbound DNA or mRNA fragment. Briefly, the target nucleotide sequence is labeled. Then the labeled target nucleotide sequence is brought into contact with either a total nuclear extract from cells containing regulation factors, or with different candidate molecules to be tested. The interaction between the target regulatory sequence of the GENSET gene and the candidate molecule or the regulation factor is detected after gel or capillary electrophoresis through a retardation in the migration.

Nucleic acids encoding proteins which are able to interact with the promoter sequence of the GENSET gene, more particularly a nucleotide sequence selected from the group consisting of the polynucleotides of the 5' and 3' regulatory region or a fragment or variant thereof, may be identified by using a one-hybrid system, such as that described in the booklet enclosed in the Matchmaker One-Hybrid System kit from Clontech (Catalog Ref. n° K1603-1), the technical teachings of which are herein incorporated by reference. Briefly, the target nucleotide sequence is cloned upstream of a selectable reporter sequence and the resulting polynucleotide construct is integrated in the yeast genome (*Saccharomyces cerevisiae*). Preferably, multiple copies of the target sequences are inserted into the reporter plasmid in tandem. The yeast cells containing the reporter sequence in their genome are then transformed with a library comprising fusion molecules between cDNAs encoding candidate proteins for binding onto the regulatory sequences of the GENSET gene and sequences encoding the activator domain of a yeast transcription factor such as GAL4. The recombinant yeast cells are plated in a culture broth for selecting cells expressing the reporter sequence. The recombinant yeast cells thus selected contain a fusion protein that is able to bind onto the target regulatory sequence of the GENSET gene. Then, the cDNAs encoding the fusion proteins are sequenced and may be cloned into expression or transcription vectors *in vitro*. The binding of the encoded polypeptides to the target regulatory sequences of the GENSET gene

may be confirmed by techniques familiar to the one skilled in the art, such as gel retardation assays or DNAse protection assays.

Ligands interacting with GENSET polypeptides

For the purpose of the present invention, a ligand means a molecule, such as a protein, a peptide, an antibody or any synthetic chemical compound capable of binding to a GENSET protein or one of its fragments or variants or to modulate the expression of the polynucleotide coding for GENSET or a fragment or variant thereof.

In the ligand screening method according to the present invention, a biological sample or a defined molecule to be tested as a putative ligand of a GENSET protein is brought into contact with the corresponding purified GENSET protein, for example the corresponding purified recombinant GENSET protein produced by a recombinant cell host as described herein, in order to form a complex between this protein and the putative ligand molecule to be tested.

As an illustrative example, to study the interaction of a GENSET protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384 as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool, with drugs or small molecules, such as molecules generated through combinatorial chemistry approaches, the microdialysis coupled to HPLC method described by Wang *et al.* (1997) or the affinity capillary electrophoresis method described by Bush *et al.* (1997), the disclosures of which are incorporated by reference, can be used.

In further methods, peptides, drugs, fatty acids, lipoproteins, or small molecules which interact with a GENSET protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool may be identified using assays such as the following. The molecule to be tested for binding is labeled with a detectable label, such as a fluorescent , radioactive, or enzymatic tag and placed in contact with immobilized GENSET protein, or a fragment thereof under conditions which permit specific binding to occur. After removal of non-specifically bound molecules, bound molecules are detected using appropriate means.

Various candidate substances or molecules can be assayed for interaction with a GENSET polypeptide. These substances or molecules include, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as polypeptides. When the candidate substance or molecule comprises a polypeptide, this polypeptide may be the resulting expression

product of a phage clone belonging to a phage-based random peptide library, or alternatively the polypeptide may be the resulting expression product of a cDNA library cloned in a vector suitable for performing a two-hybrid screening assay.

A. Candidate ligands obtained from random peptide libraries

In a particular embodiment of the screening method, the putative ligand is the expression product of a DNA insert contained in a phage vector (Parmley and Smith, 1988). Specifically, random peptide phages libraries are used. The random DNA inserts encode for peptides of 8 to 20 amino acids in length (Oldenburg *et al.*, 1992; Valadon *et al.*, 1996; Lucas, 1994; Westerink, 1995; Felici *et al.*, 1991), which disclosures are hereby incorporated by reference in their entireties. According to this particular embodiment, the recombinant phages expressing a protein that binds to an immobilized GENSET protein is retained and the complex formed between the GENSET protein and the recombinant phage may be subsequently immunoprecipitated by a polyclonal or a monoclonal antibody directed against the GENSET protein.

Once the ligand library in recombinant phages has been constructed, the phage population is brought into contact with the immobilized GENSET protein. Then the preparation of complexes is washed in order to remove the non-specifically bound recombinant phages. The phages that bind specifically to the GENSET protein are then eluted by a buffer (acid pH) or immunoprecipitated by the monoclonal antibody produced by the hybridoma anti-GENSET, and this phage population is subsequently amplified by an over-infection of bacteria (for example E. coli). The selection step may be repeated several times, preferably 2-4 times, in order to select the more specific recombinant phage clones. The last step comprises characterizing the peptide produced by the selected recombinant phage clones either by expression in infected bacteria and isolation, expressing the phage insert in another host-vector system, or sequencing the insert contained in the selected recombinant phages.

B. Candidate ligands obtained by competition experiments.

Alternatively, peptides, drugs or small molecules which bind to a GENSET protein or fragment thereof comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool, may be identified in competition experiments. In such assays, the GENSET protein, or a fragment thereof, is immobilized to a surface, such as a plastic plate. Increasing amounts of the peptides, drugs or small molecules are placed in contact with the immobilized GENSET protein, or a fragment thereof, in the presence of a detectable labeled known GENSET protein ligand. For example, the GENSET ligand may be detectably labeled with a fluorescent, radioactive, or enzymatic tag. The ability of the test molecule

to bind the GENSET protein, or a fragment thereof, is determined by measuring the amount of detectably labeled known ligand bound in the presence of the test molecule. A decrease in the amount of known ligand bound to the GENSET protein, or a fragment thereof, when the test molecule is present indicated that the test molecule is able to bind to the GENSET protein, or a fragment thereof.

C. Candidate ligands obtained by affinity chromatography.

Proteins or other molecules interacting with a GENSET protein, or a fragment thereof comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool, can also be found using affinity columns which contain the GENSET protein, or a fragment thereof. The GENSET protein, or a fragment thereof, may be attached to the column using conventional techniques including chemical coupling to a suitable column matrix such as agarose, Affi Gel® , or other matrices familiar to those of skill in art. In some embodiments of this method, the affinity column contains chimeric proteins in which the GENSET protein, or a fragment thereof, is fused to glutathion S transferase (GST). A mixture of cellular proteins or pool of expressed proteins as described above is applied to the affinity column. Proteins or other molecules interacting with the GENSET protein, or a fragment thereof, attached to the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen *et al.* (1997), the disclosure of which is incorporated by reference. Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

D. Candidate ligands obtained by optical biosensor methods

Proteins interacting with a GENSET protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool, can also be screened by using an Optical Biosensor as described in Edwards and Leatherbarrow (1997) and also in Szabo *et al.* (1995), the disclosures of which are incorporated by reference. This technique permits the detection of interactions between molecules in real time, without the need of labeled molecules. This technique is based on the surface plasmon resonance (SPR) phenomenon. Briefly, the candidate ligand molecule to be tested is attached to a surface (such as a carboxymethyl dextran matrix). A light beam is directed towards the side of the surface

that does not contain the sample to be tested and is reflected by said surface. The SPR phenomenon causes a decrease in the intensity of the reflected light with a specific association of angle and wavelength. The binding of candidate ligand molecules cause a change in the refraction index on the surface, which change is detected as a change in the SPR signal. For screening of candidate ligand molecules or substances that are able to interact with the GENSET protein, or a fragment thereof, the GENSET protein, or a fragment thereof, is immobilized onto a surface. This surface comprises one side of a cell through which flows the candidate molecule to be assayed. The binding of the candidate molecule on the GENSET protein, or a fragment thereof, is detected as a change of the SPR signal. The candidate molecules tested may be proteins, peptides, carbohydrates, lipids, or small molecules generated by combinatorial chemistry. This technique may also be performed by immobilizing eukaryotic or prokaryotic cells or lipid vesicles exhibiting an endogenous or a recombinantly expressed GENSET protein at their surface.

The main advantage of the method is that it allows the determination of the association rate between the GENSET protein and molecules interacting with the GENSET protein. It is thus possible to select specifically ligand molecules interacting with the GENSET protein, or a fragment thereof, through strong or conversely weak association constants.

E. Candidate ligands obtained through a two-hybrid screening assay.

The yeast two-hybrid system is designed to study protein-protein interactions *in vivo* (Fields and Song, 1989), which disclosure is hereby incorporated by reference in its entirety, and relies upon the fusion of a bait protein to the DNA binding domain of the yeast Gal4 protein. This technique is also described in the US Patent N° US 5,667,973 and the US Patent N° 5,283,173, the technical teachings of both patents being herein incorporated by reference.

The general procedure of library screening by the two-hybrid assay may be performed as described by Harper *et al.* (1993) or as described by Cho *et al.* (1998) or also Fromont-Racine *et al.* (1997), which disclosures are hereby incorporated by reference in their entireties.

The bait protein or polypeptide comprises, consists essentially of, or consists of a GENSET polypeptide or a fragment thereof comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool.

More precisely, the nucleotide sequence encoding the GENSET polypeptide or a fragment or variant thereof is fused to a polynucleotide encoding the DNA binding domain of the GAL4 protein, the fused nucleotide sequence being inserted in a suitable expression vector, for example pAS2 or pM3.

Then, a human cDNA library is constructed in a specially designed vector, such that the human cDNA insert is fused to a nucleotide sequence in the vector that encodes the transcriptional domain of the GAL4 protein. Preferably, the vector used is the pACT vector. The polypeptides encoded by the nucleotide inserts of the human cDNA library are termed "pray" polypeptides.

A third vector contains a detectable marker gene, such as beta galactosidase gene or CAT gene that is placed under the control of a regulation sequence that is responsive to the binding of a complete Gal4 protein containing both the transcriptional activation domain and the DNA binding domain. For example, the vector pG5EC may be used.

Two different yeast strains are also used. As an illustrative but non limiting example the two different yeast strains may be the followings :

- Y190, the phenotype of which is (MATa, Leu2-3, 112 ura3-12, trp1-901, his3-D200, ade2-101, gal4Dgal180D URA3 GAL-LacZ, LYS GAL-HIS3, cyh^r);

- Y187, the phenotype of which is (MATa gal4 gal80 his3 trp1-901 ade2-101 ura3-52 leu2-3, -112 URA3 GAL-lacZmet^-), which is the opposite mating type of Y190.

Briefly, 20 µg of pAS2/GENSET and 20 µg of pACT-cDNA library are co-transformed into yeast strain Y190. The transformants are selected for growth on minimal media lacking histidine, leucine and tryptophan, but containing the histidine synthesis inhibitor 3-AT (50 mM). Positive colonies are screened for beta galactosidase by filter lift assay. The double positive colonies (His^+, beta-gal^+) are then grown on plates lacking histidine, leucine, but containing tryptophan and cycloheximide (10 mg/ml) to select for loss of pAS2/GENSET plasmids but retention of pACT-cDNA library plasmids. The resulting Y190 strains are mated with Y187 strains expressing GENSET or non-related control proteins; such as cyclophilin B, lamin, or SNF1, as Gal4 fusions as described by Harper *et al.* (1993) and by Bram *et al.* (1993), which disclosures are hereby incorporated by reference in their entireties, and screened for beta galactosidase by filter lift assay. Yeast clones that are beta gal- after mating with the control Gal4 fusions are considered false positives.

In another embodiment of the two-hybrid method according to the invention, interaction between the GENSET or a fragment or variant thereof with cellular proteins may be assessed using the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual accompanying the kit, the disclosure of which is incorporated herein by reference, nucleic acids encoding the GENSET protein or a portion thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. A desired cDNA, preferably human cDNA, is inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking

histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain interaction between GENSET and the protein or peptide encoded by the initially selected cDNA insert.

Compounds Modulating GENSET biological activity

Another method of screening for compounds that modulate GENSET gene expression and/or biological activity is by measuring the effects of test compounds on a given cellular property in a host cell, such as apoptosis, proliferation, differentiation, protein glycosylation, etc... using a variety of techniques known to those skilled in the art including those described herein and especially in the section entitled "**Erreur! Source du renvoi introuvable.**".

In one embodiment, the present invention relates to a method of identifying an agent which alters GENSET activity, wherein a nucleic acid construct comprising a nucleic acid which encodes a mammalian GENSET polypeptide is introduced into a host cell. The host cells produced are maintained under conditions appropriate for expression of the encoded mammalian GENSET polypeptides, whereby the nucleic acid is expressed. The host cells are then contacted with a compound to be assessed (an agent) and the given cellular property of the cells is detected in the presence of the compound to be assessed. Detection of a change in the given cellular property in the presence of the agent indicates that the agent alters GENSET activity.

In a particular embodiment, the invention relates to a method of identifying an agent which is an activator of GENSET activity, wherein detection of a change of the given cellular property in the presence of the agent indicates that the agent activates GENSET activity. In another particular embodiment, the invention relates to a method of identifying an agent which is an inhibitor of GENSET activity, wherein detection of a change of the given cellular property in the presence of the agent indicates that the agent inhibits GENSET activity.

Methods of Screening for Compounds Modulating GENSET Expression and/or Activity

The present invention also relates to methods of screening compounds for their ability to modulate (e.g. increase or inhibit) the activity or expression of GENSET. More specifically, the present invention relates to methods of testing compounds for their ability either to increase or to decrease expression or activity of GENSET. The assays are performed *in vitro* or *in vivo.*

In vitro methods

In vitro, cells expressing GENSET are incubated in the presence and absence of the test compound. By determining the level of GENSET expression in the presence of the test compound or the level of GENSET activity in the presence of the test compound, compounds can be identified that suppress or enhance GENSET expression or activity. Alternatively, constructs comprising a GENSET regulatory sequence operably linked to a reporter gene (e.g. luciferase, chloramphenicol acetyl transferase, LacZ, green fluorescent protein, etc.) can be introduced into host cells and the

effect of the test compounds on expression of the reporter gene detected. Cells suitable for use in the foregoing assays include, but are not limited to, cells having the same origin as tissues or cell lines in which the polypeptide is known to be expressed using the data from Table IX.

Consequently, the present invention encompasses a method for screening molecules that modulate the expression of a GENSET gene, said screening method comprising the steps of:

a) cultivating a prokaryotic or an eukaryotic cell that has been transfected with a nucleotide sequence encoding a GENSET protein or a variant or a fragment thereof, placed under the control of its own promoter;

b) bringing into contact said cultivated cell with a molecule to be tested;

c) quantifying the expression of said GENSET protein or a variant or a fragment thereof in the presence of said molecule.

Using DNA recombination techniques well known by the one skill in the art, the GENSET protein encoding DNA sequence is inserted into an expression vector, downstream from its promoter sequence. As an illustrative example, the promoter sequence of the GENSET gene is contained in the 5' untranscribed region of the GENSET genomic DNA.

The quantification of the expression of a GENSET protein may be realized either at the mRNA level (using for example Northen blots, RT-PCR, preferably quantitative RT-PCR with primers and probes specific for the GENSET mRNA of interest) or at the protein level (using polyclonal or monoclonal antibodies in immunoassays such as ELISA or RIA assays, Western blots, or immunochemistry).

The present invention also concerns a method for screening substances or molecules that are able to increase, or in contrast to decrease, the level of expression of a GENSET gene. Such a method may allow the one skilled in the art to select substances exerting a regulating effect on the expression level of a GENSET gene and which may be useful as active ingredients included in pharmaceutical compositions for treating patients suffering from disorders associated with abnormal levels of GENSET products.

Thus, also part of the present invention is a method for screening a candidate molecule that modulates the expression of a GENSET gene, this method comprises the following steps:

a) providing a recombinant cell host containing a nucleic acid, wherein said nucleic acid comprises a GENSET 5' regulatory region or a regulatory active fragment or variant thereof, operably linked to a polynucleotide encoding a detectable protein;

b) obtaining a candidate molecule; and

c) determining the ability of said candidate molecule to modulate the expression levels of said polynucleotide encoding the detectable protein.

In a further embodiment, said nucleic acid comprising a GENSET 5' regulatory region or a regulatory active fragment or variant thereof, includes the 5'UTR region of a GENSET cDNA selected from the group comprising of the 5'UTRs of the sequences of SEQ ID Nos 1-241,

sequences of clones inserts of the deposited clone pool, regulatory active fragments and variants thereof. In a more preferred embodiment of the above screening method, said nucleic acid includes a promoter sequence which is endogenous with respect to the GENSET 5'UTR sequence. In another more preferred embodiment of the above screening method, said nucleic acid includes a promoter sequence which is exogenous with respect to the GENSET 5'UTR sequence defined therein.

Preferred polynucleotides encoding a detectable protein are polynucleotides encoding beta galactosidase, green fluorescent protein (GFP) and chloramphenicol acetyl transferase (CAT).

The invention further relates to a method for the production of a pharmaceutical composition comprising a method of screening a candidate molecule that modulates the expression of a GENSET gene and furthermore mixing the identified molecule with a pharmaceutically acceptable carrier.

The invention also pertains to kits for the screening of a candidate substance modulating the expression of a GENSET gene. Preferably, such kits comprise a recombinant vector that allows the expression of a GENSET 5' regulatory region or a regulatory active fragment or a variant thereof, operably linked to a polynucleotide encoding a detectable protein or a GENSET protein or a fragment or a variant thereof. More preferably, such kits include a recombinant vector that comprises a nucleic acid including the 5'UTR region of a GENSET cDNA selected from the group comprising the 5'UTRs of the sequences of SEQ ID Nos 1-241, sequences of clones inserts of the deposited clone pool, regulatory active fragments and variants thereof, being operably linked to a polynucleotide encoding a detectable protein.

For the design of suitable recombinant vectors useful for performing the screening methods described above, it will be referred to the section of the present specification wherein the preferred recombinant vectors of the invention are detailed.

Another object of the present invention comprises methods and kits for the screening of candidate substances that interact with a GENSET polypeptide, fragments or variants thereof. By their capacity to bind covalently or non-covalently to a GENSET protein, fragments or variants thereof, these substances or molecules may be advantageously used both *in vitro* and *in vivo*.

In vitro, said interacting molecules may be used as detection means in order to identify the presence of a GENSET protein in a sample, preferably a biological sample.

A method for the screening of a candidate substance that interact with a GENSET polypeptide, fragments or variants thereof, said methods comprising the following steps:

a) providing a polypeptide comprising, consisting essentially of, or consisting of a GENSET protein or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides

included in SEQ ID Nos: 242-272 and 274-384 as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool;

b) obtaining a candidate substance;

c) bringing into contact said polypeptide with said candidate substance;

d) detecting the complexes formed between said polypeptide and said candidate substance.

The invention further relates to a method for the production of a pharmaceutical composition comprising a method for the screening of a candidate substance that interact with a GENSET polypeptide, fragments or variants thereof and furthermore mixing the identified substance with a pharmaceutically acceptable carrier.

The invention further concerns a kit for the screening of a candidate substance interacting with the GENSET polypeptide, wherein said kit comprises:

a) a polypeptide comprising, consisting essentially of, or consisting of a GENSET protein or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384 as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool; and

b) optionally means useful to detect the complex formed between said polypeptide or a variant thereof and the candidate substance.

In a preferred embodiment of the kit described above, the detection means comprises a monoclonal or polyclonal antibody binding to said GENSET protein or fragment or variant thereof.

In vivo methods

Compounds that suppress or enhance GENSET expression can also be identified using *in vivo* screens. In these assays, the test compound is administered (e.g. IV, IP, IM, orally, or otherwise), to the animal, for example, at a variety of dose levels. The effect of the compound on GENSET expression is determined by comparing GENSET levels, for example in tissues known to express the gene of interest using, for example the data obtained in Table IX, and using Northern blots, immunoassays, PCR, etc., as described above. Suitable test animals include rodents (e.g., mice and rats), primates, mammals. Humanized mice can also be used as test animals, that is mice in which the endogenous mouse protein is ablated (knocked out) and the homologous human protein added back by standard transgenic approaches. Such mice express only the human form of a protein. Humanized mice expressing only the human GENSET can be used to study *in vivo* responses to potential agents regulating GENSET protein or mRNA levels. As an example, transgenic mice have been produced carrying the human apoE4 gene. They are then bred with a mouse line that lacks endogenous apoE, to produce an animal model carrying human proteins believed to be instrumental in development of Alzheimer's pathology. Such transgenic animals are useful for dissecting the biochemical and physiological steps of disease, and for development of

therapies for disease intervention (Loring, *et al*, 1996) (incorporated herein by reference in its entirety).

Uses for compounds modulating GENSET expression and/or biological activity

Using *in vivo* (or *in vitro)* systems, it may be possible to identify compounds that exert a tissue specific effect, for example, that increase GENSET expression or activity only in tissues of interest. Screening procedures such as those described above are also useful for identifying agents for their potential use in pharmacological intervention strategies. Agents that enhance GENSET expression or stimulate its activity may thus be used to treat disorders which require upregulated levels of GENSET gene expression and/or activity. Compounds that suppress GENSET expression or inhibit its activity can be used to treat disorders which require downregulated levels of GENSET gene expression and/or activity.

Also encompassed by the present invention is an agent which interacts with GENSET directly or indirectly, and inhibits or enhances GENSET expression and/or function. In one embodiment, the agent is an inhibitor which interferes with GENSET directly (e.g., by binding GENSET) or indirectly (e.g., by blocking the ability of GENSET to have a GENSET biological activity). In a particular embodiment, an inhibitor of GENSET protein is an antibody specific for GENSET protein or a functional portion of GENSET; that is, the antibody binds a GENSET polypeptide. For example, the antibody can be specific for a polypeptide encoded by one of the amino acid sequences of human GENSET genes (SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool), mammal GENSET or portions thereof. Alternatively, the inhibitor can be an agent other than an antibody (e.g., small organic molecule, protein or peptide) which binds GENSET and blocks its activity. For example, the inhibitor can be an agent which mimics GENSET structurally, but lacks its function. Alternatively, it can be an agent which binds to or interacts with a molecule which GENSET normally binds with or interacts with, thus blocking GENSET from doing so and preventing it from exerting the effects it would normally exert.

In another embodiment, the agent is an enhancer (activator) of GENSET which increases the activity of GENSET (increases the effect of a given amount or level of GENSET), increases the length of time it is effective (by preventing its degradation or otherwise prolonging the time during which it is active) or both either directly or indirectly.

The GENSET sequences of the present invention can also be used to generate nonhuman gene knockout animals, such as mice, which lack a GENSET gene or transgenically overexpress GENSET. For example, such GENSET gene knockout mice can be generated and used to obtain further insight into the function of GENSET as well as assess the specificity of GENSET activators and inhibitors. Also, over expression of GENSET (e.g., human GENSET) in transgenic mice can be used as a means of creating a test system for GENSET activators and inhibitors (e.g., against

human GENSET). In addition, the GENSET gene can be used to clone the GENSET promoter/enhancer in order to identify regulators of GENSET transcription. GENSET gene knockout animals include animals which completely or partially lack the GENSET gene and/or GENSET activity or function. Thus the present invention relates to a method of inhibiting (partially or completely) GENSET biological activty in a mammal (e.g., human) comprising administering to the mammal an effective amount of an inhibitor of GENSET. The invention also relates to a method of enhancing GENSET biological activity in a mammal comprising administering to the mammal an effective amount of an enhancer GENSET.

Inhibiting GENSET expression

Therapeutic compositions according to the present invention may comprise advantageously one or several GENSET oligonucleotide fragments as an antisense tool or a triple helix tool that inhibits the expression of the corresponding GENSET gene.

Antisense Approach

In antisense approaches, nucleic acid sequences complementary to an mRNA are hybridized to the mRNA intracellularly, thereby blocking the expression of the protein encoded by the mRNA. The antisense nucleic acid molecules to be used in gene therapy may be either DNA or RNA sequences. Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel *et al.*(1995), which disclosure is hereby incorporated by reference in its entirety.

Preferably, the antisense tools are chosen among the polynucleotides (15-200 bp long) that are complementary to GENSET mRNA, more preferably to the 5'end of the GENSET mRNA. In another embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targeted gene are used.

Other preferred antisense polynucleotides according to the present invention are sequences complementary to either a sequence of GENSET mRNAs comprising the translation initiation codon ATG or a sequence of GENSET genomic DNA containing a splicing donor or acceptor site.

Preferably, the antisense polynucleotides of the invention have a 3' polyadenylation signal that has been replaced with a self-cleaving ribozyme sequence, such that RNA polymerase II transcripts are produced without poly(A) at their 3' ends, these antisense polynucleotides being incapable of export from the nucleus, such as described by Liu *et al.*(1994), which disclosure is hereby incorporated by reference in its entirety. In a preferred embodiment, these GENSET antisense polynucleotides also comprise, within the ribozyme cassette, a histone stem-loop structure to stabilize cleaved transcripts against 3'-5' exonucleolytic degradation, such as the structure described by Eckner *et al.*(1991), which disclosure is hereby incorporated by reference in its entirety.

The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex having sufficient stability to inhibit the expression of the GENSET mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green *et al.*, (1986) and Izant and Weintraub, (1984), the disclosures of which are incorporated herein by reference.

In some strategies, antisense molecules are obtained by reversing the orientation of the GENSET coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using *in vitro* transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach involves transcription of GENSET antisense nucleic acids *in vivo* by operably linking DNA containing the antisense sequence to a promoter in a suitable expression vector.

Alternatively, oligonucleotides which are complementary to the strand normally transcribed in the cell may be synthesized *in vitro*. Thus, the antisense nucleic acids are complementary to the corresponding mRNA and are capable of hybridizing to the mRNA to create a duplex. In some embodiments, the antisense sequences may contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity. Examples of modifications suitable for use in antisense strategies include 2' O-methyl RNA oligonucleotides and Protein-nucleic acid (PNA) oligonucleotides. Further examples are described by Rossi *et al.*, (1991), which disclosure is hereby incorporated by reference in its entirety.

Various types of antisense oligonucleotides complementary to the sequence of the GENSET cDNA or genomic DNA may be used. In one preferred embodiment, stable and semi-stable antisense oligonucleotides described in International Application No. PCT WO94/23026, hereby incorporated by reference, are used. In these molecules, the 3' end or both the 3' and 5' ends are engaged in intramolecular hydrogen bonding between complementary base pairs. These molecules are better able to withstand exonuclease attacks and exhibit increased stability compared to conventional antisense oligonucleotides.

In another preferred embodiment, the antisense oligodeoxynucleotides against herpes simplex virus types 1 and 2 described in International Application No. WO 95/04141, hereby incorporated by reference, are used.

In yet another preferred embodiment, the covalently cross-linked antisense oligonucleotides described in International Application No. WO 96/31523, hereby incorporated by reference, are used. These double- or single-stranded oligonucleotides comprise one or more, respectively, inter- or intra-oligonucleotide covalent cross-linkages, wherein the linkage consists of an amide bond between a primary amine group of one strand and a carboxyl group of the other strand or of the same strand, respectively, the primary amine group being directly substituted in the 2' position of the strand nucleotide monosaccharide ring, and the carboxyl group being carried by an aliphatic

spacer group substituted on a nucleotide or nucleotide analog of the other strand or the same strand, respectively.

The antisense oligodeoxynucleotides and oligonucleotides disclosed in International Application No. WO 92/18522, incorporated by reference, may also be used. These molecules are stable to degradation and contain at least one transcription control recognition sequence which binds to control proteins and are effective as decoys therefor. These molecules may contain "hairpin" structures, "dumbbell" structures, "modified dumbbell" structures, "cross-linked" decoy structures and "loop" structures.

In another preferred embodiment, the cyclic double-stranded oligonucleotides described in European Patent Application No. 0 572 287 A2, hereby incorporated by reference are used. These ligated oligonucleotide "dumbbells" contain the binding site for a transcription factor and inhibit expression of the gene under control of the transcription factor by sequestering the factor.

Use of the closed antisense oligonucleotides disclosed in International Application No. WO 92/19732, hereby incorporated by reference, is also contemplated. Because these molecules have no free ends, they are more resistant to degradation by exonucleases than are conventional oligonucleotides. These oligonucleotides may be multifunctional, interacting with several regions which are not adjacent to the target mRNA.

The appropriate level of antisense nucleic acids required to inhibit gene expression may be determined using *in vitro* expression analysis. The antisense molecule may be introduced into the cells by diffusion, injection, infection or transfection using procedures known in the art. For example, the antisense nucleic acids can be introduced into the body as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsidated by viral protein, or as an oligonucleotide operably linked to a promoter contained in an expression vector. The expression vector may be any of a variety of expression vectors known in the art, including retroviral or viral vectors, vectors capable of extrachromosomal replication, or integrating vectors. The vectors may be DNA or RNA.

The antisense molecules are introduced onto cell samples at a number of different concentrations preferably between $1x10^{-10}$M to $1x10^{-4}$M. Once the minimum concentration that can adequately control gene expression is identified, the optimized dose is translated into a dosage suitable for use *in vivo*. For example, an inhibiting concentration in culture of $1x10^{-7}$ translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of oligonucleotide approaching 100 mg/kg bodyweight or higher may be possible after testing the toxicity of the oligonucleotide in laboratory animals. It is additionally contemplated that cells from the vertebrate are removed, treated with the antisense oligonucleotide, and reintroduced into the vertebrate.

In a preferred application of this invention, the polypeptide encoded by the gene is first identified, so that the effectiveness of antisense inhibition on translation can be monitored using

techniques that include but are not limited to antibody-mediated tests such as RIAs and ELISA, functional assays, or radiolabeling.

An alternative to the antisense technology that is used according to the present invention comprises using ribozymes that will bind to a target sequence via their complementary polynucleotide tail and that will cleave the corresponding RNA by hydrolyzing its target site (namely "hammerhead ribozymes"). Briefly, the simplified cycle of a hammerhead ribozyme comprises (1) sequence specific binding to the target RNA via complementary antisense sequences; (2) site-specific hydrolysis of the cleavable motif of the target strand; and (3) release of cleavage products, which gives rise to another catalytic cycle. Indeed, the use of long-chain antisense polynucleotide (at least 30 bases long) or ribozymes with long antisense arms are advantageous. A preferred delivery system for antisense ribozyme is achieved by covalently linking these antisense ribozymes to lipophilic groups or to use liposomes as a convenient vector. Preferred antisense ribozymes according to the present invention are prepared as described by Rossi *et al*, (1991) and Sczakiel *et al.*(1995), the specific preparation procedures being referred to in said articles being herein incorporated by reference.

Triple Helix Approach

The GENSET genomic DNA may also be used to inhibit the expression of the GENSET gene based on intracellular triple helix formation.

Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity when it is associated with a particular gene. The GENSET cDNAs or genomic DNAs of the present invention or, more preferably, a fragment of those sequences, can be used to inhibit gene expression in individuals having diseases associated with expression of a particular gene. Similarly, a portion of the GENSET genomic DNA can be used to study the effect of inhibiting GENSET transcription within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. Thus, both types of sequences from the GENSET genomic DNA are contemplated within the scope of this invention.

To carry out gene therapy strategies using the triple helix approach, the sequences of the GENSET genomic DNA are first scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting GENSET expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting GENSET expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which express the GENSET gene.

The oligonucleotides can be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced GENSET expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the GENSET gene in cells which have been treated with the oligonucleotide. The cell functions to be monitored are predicted based upon the homologies of the target gene corresponding to the cDNA from which the oligonucleotide was derived with known gene sequences that have been associated with a particular function. The cell functions can also be predicted based on the presence of abnormal physiology within cells derived from individuals with a particular inherited disease, particularly when the cDNA is associated with the disease using techniques described in the section entitled "Identification of genes associated with hereditary diseases or drug response".

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced *in vivo* using the techniques and at a dosage calculated based on the *in vitro* results, as described in the section entitled "Antisense Approach".

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin *et al.*(1989), which is hereby incorporated by this reference.

Treating GENSET-related disorders

The present invention further relates to methods of treating diseases/disorders by increasing GENSET activity and/or expression. The invention also relates to methods of treating diseases/disorders by decreasing GENSET activity and or expression. These methodologies can be effected using compounds selected using screening protocols such as those described herein and/or by using the gene therapy and antisense approaches described in the art and herein. Gene therapy can be used to effect targeted expression of GENSET. The GENSET coding sequence can be cloned into an appropriate expression vector and targeted to a particular cell type(s) to achieve efficient, high level expression. Introduction of the GENSET coding sequence into target cells can be achieved, for example, using particle mediated DNA delivery, (Haynes, 1996 and Maurer, 1999), direct injection of naked DNA, (Levy *et al.*, 1996; and Felgner, 1996), or viral vector mediated transport (Smith *et al.*, 1996, Stone *et al*, 2000; Wu and Atai, 2000), each of which disclosures are hereby incorporated by reference in their entireties . Tissue specific effects can be achieved, for

example, in the case of virus mediated transport by using viral vectors that are tissue specific, or by the use of promoters that are tissue specific.

Combinatorial approaches can also be used to ensure that the GENSET coding sequence is activated in the target tissue (Butt and Karathanasis, 1995; Miller and Whelan, 1997), which disclosures are hereby incorporated by reference in their entireties. Antisense oligonucleotides complementary to GENSET mRNA can be used to selectively diminish or ablate the expression of the protein, for example, at sites of inflammation. More specifically, antisense constructs or antisense oligonucleotides can be used to inhibit the production of GENSET in high expressing cells such as those cited in the third column of Table X. Antisense mRNA can be produced by transfecting into target cells an expression vector with the GENSET gene sequence, or portion thereof, oriented in an antisense direction relative to the direction of transcription. Appropriate vectors include viral vectors, including retroviral, adenoviral, and adeno-associated viral vectors, as well as nonviral vectors. Tissue specific promoters can be used. Alternatively, antisense oligonucleotides can be introduced directly into target cells to achieve the same goal. (See also other delivery methodologies described herein in connection with gene therapy.). Oligonucleotides can be selected/designed to achieve a high level of specificity (Wagner *et al.*, 1996), which disclosure is hereby incorporated by reference in its entirety. The therapeutic methodologies described herein are applicable to both human and non-human mammals (including cats and dogs).

PHARMACEUTICAL AND PHYSIOLOGICALLY ACCEPTABLE COMPOSITIONS

The present invention also relates to pharmaceutical or physiologically acceptable compositions comprising, as active agent, the polypeptides, nucleic acids or antibodies of the invention. The invention also relates to compositions comprising, as active agent, compounds selected using the above-described screening protocols. Such compositions include the active agent in combination with a pharmaceutical or physiologically acceptably acceptable carrier. In the case of naked DNA, the "carrier" may be gold particles. The amount of active agent in the composition can vary with the agent, the patient and the effect sought. Likewise, the dosing regimen can vary depending on the composition and the disease/disorder to be treated.

Therefore, the invention related to methods for the production of pharmaceutical composition comprising a method for selecting an active agent, compound, substance or molecule using any of the screening method described herein and furthermore mixing the identified active agent, compound, substance or molecule with a pharmaceutically acceptable carrier.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means. In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising

excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack PublishingCo. Easton, Pa).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through a combination of active compounds with solid excipient, sulting mixture is optionally grinding, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titaniumdioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquidpolyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethylcellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes.

Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of GENSET, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example GENSET or fragments thereof, antibodies of GENSET, agonists, antagonists or inhibitors of GENSET, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions maybe administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

USES OF GENSET SEQUENCES: COMPUTER-RELATED EMBODIMENTS

As used herein the term "cDNA codes of SEQ ID Nos: 1-241" encompasses the nucleotide sequences of SEQ ID Nos: 1-241 and of clones inserts of the deposited clone pool, fragments thereof, nucleotide sequences homologous thereto, and sequences complementary to all of the preceding sequences. The fragments include fragments of SEQ ID Nos: 1-241 comprising at least 8, 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 1000 or 2000 consecutive nucleotides of SEQ ID Nos: 1-241. Preferably the fragments include signal sequences and coding sequences for mature polypeptides of SEQ ID Nos: 1-31 and 33-143, polynucleotides described in Tables Va and Table Vb, polynucleotides encoding polypeptides described in Table VI, polynucleotide described herein as encoding polypeptides having a biological activity, or fragments comprising at least 8, 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 1000 or 2000 consecutive nucleotides of the signal sequences or coding sequences for mature polypeptides of SEQ ID Nos: 1-31 and 33-143, polynucleotides described in Tables Va and Table Vb, polynucleotides encoding polypeptides described in Table VI, and polynucleotide described herein as encoding polypeptides having a biological activity. Homologous sequences and fragments of SEQ ID Nos: 1-241 refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, or 75% identity to these sequences. Identity may be determined using any of the computer programs and parameters described herein, including BLAST2N with the default parameters or with any modified parameters. Homologous sequences also include RNA sequences in which uridines replace the thymines in the cDNA codes of SEQ ID Nos: 1-241. The homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error as described above. Preferably the homologous sequences and fragments of SEQ

ID Nos: 1-241 include polynucleotides homologous to signal sequences and coding sequences for mature polypeptides of SEQ ID Nos: 1-31 and 33-143, polynucleotides described in Tables Va and Table Vb, polynucleotides encoding a polypeptide fragment described as a domain in Table VI, polynucleotide described herein as encoding polypeptides having a biological activity, or fragments comprising at least 8, 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 1000 or 2000 consecutive nucleotides of the signal sequences and coding sequences for mature polypeptides of SEQ ID Nos: 1-31 and 33-143, polynucleotides described in Tables Va and Table Vb, polynucleotides described in Table VI, and polynucleotide described herein as encoding polypeptides having a biological activity. It will be appreciated that the cDNA codes of SEQ ID Nos: 1-241 can be represented in the traditional single character format (See the inside back cover of Styer, 1995) or in any other format which records the identity of the nucleotides in a sequence.

As used herein the term "polypeptide codes of SEQ ID Nos: 242-482" encompasses the polypeptide sequences of SEQ ID Nos: 242-482, the signal peptides included in SEQ ID Nos: 242-272 and 274-384, the mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, the full-length, signal peptides and mature polypeptide sequences encoded by the clone inserts of the deposited clone pool, polypeptide sequences homologous thereto, or fragments of any of the preceding sequences. Homologous polypeptide sequences refer to a polypeptide sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75% identity to one of the polypeptide sequences of SEQ ID Nos: 242-482, the signal peptides included in SEQ ID Nos: 242-272 and 274-384, the mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, the full-length, signal peptides and mature polypeptide sequences encoded by the clone inserts of the deposited clone pool. Identity may be determined using any of the computer programs and parameters described herein, including FASTA with the default parameters or with any modified parameters. The homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error as described above. The polypeptide fragments comprise at least 5, 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, 200, 250, 300, 350, 400, 450 or 500 consecutive amino acids of the polypeptides of SEQ ID Nos: 242-482. Preferably, the fragments include polypeptides encoded by the signal peptides included in SEQ ID Nos: 242-272 and 274-384, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, polynucleotides described in Tables Va and in Table Vb, domains described in Table VI, epitopes described in Table VII, polypeptides described herein as having a biological activity, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300 or 400 consecutive amino acids of the signal peptides included in SEQ ID Nos: 242-272 and 274-384, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, the polypeptides encoded by the polynucleotides described in Tables Va and in Table Vb, domains of Table VI, epitopes of Table VII or of polypeptides described herein as having a biological activity. It will be appreciated that the polypeptide codes of the SEQ ID Nos: 242-482 can be represented in the traditional single character format or three letter

format (See the inside back cover of Stryer, 1995) or in any other format which relates the identity of the polypeptides in a sequence.

It will be appreciated by those skilled in the art that the nucleic acid codes of the invention and polypeptide codes of the invention can be stored, recorded, and manipulated on any medium which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid codes of the invention, or one or more of the polypeptide codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, 50, 75, 100, 150 or 200 nucleic acid codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, 50, 75, 100, 150 or 200 polypeptide codes of the invention.

Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random Access Memory (RAM), or Read Only Memory (ROM) as well as other types of other media known to those skilled in the art.

Embodiments of the present invention include systems, particularly computer systems which store and manipulate the sequence information described herein. One example of a computer system 100 is illustrated in block diagram form in Figure 2. As used herein, "a computer system" refers to the hardware components, software components, and data storage components used to analyze the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. In one embodiment, the computer system 100 is a Sun Enterprise 1000 server (Sun Microsystems, Palo Alto, CA). The computer system 100 preferably includes a processor for processing, accessing and manipulating the sequence data. The processor 105 can be any well-known type of central processing unit, such as the Pentium III from Intel Corporation, or similar processor from Sun, Motorola, Compaq or International Business Machines.

Preferably, the computer system 100 is a general purpose system that comprises the processor 105 and one or more internal data storage components 110 for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently available computer systems are suitable.

In one particular embodiment, the computer system 100 includes a processor 105 connected to a bus which is connected to a main memory 115 (preferably implemented as RAM) and one or more internal data storage devices 110, such as a hard drive and/or other computer readable media

having data recorded thereon. In some embodiments, the computer system 100 further includes one or more data retrieving device 118 for reading the data stored on the internal data storage devices 110.

The data retrieving device 118 may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, etc. In some embodiments, the internal data storage device 110 is a removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, etc. containing control logic and/or data recorded thereon. The computer system 100 may advantageously include or be programmed by appropriate software for reading the control logic and/or the data from the data storage component once inserted in the data retrieving device.

The computer system 100 includes a display 120 which is used to display output to a computer user. It should also be noted that the computer system 100 can be linked to other computer systems 125a-c in a network or wide area network to provide centralized access to the computer system 100.

Software for accessing and processing the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention (such as search tools, compare tools, and modeling tools etc.) may reside in main memory 115 during execution.

In some embodiments, the computer system 100 may further comprise a sequence comparer for comparing the above-described nucleic acid codes of the invention or the polypeptide codes of the invention stored on a computer readable medium to reference nucleotide or polypeptide sequences stored on a computer readable medium. A "sequence comparer" refers to one or more programs which are implemented on the computer system 100 to compare a nucleotide or polypeptide sequence with other nucleotide or polypeptide sequences and/or compounds including but not limited to peptides, peptidomimetics, and chemicals stored within the data storage means. For example, the sequence comparer may compare the nucleotide sequences of nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention stored on a computer readable medium to reference sequences stored on a computer readable medium to identify homologies, motifs implicated in biological function, or structural motifs. The various sequence comparer programs identified elsewhere in this patent specification are particularly contemplated for use in this aspect of the invention.

Figure 3 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database. The database of sequences can be a private database stored within the computer system 100, or a public database such as GENBANK, PIR OR SWISSPROT that is available through the Internet.

The process 200 begins at a start state 201 and then moves to a state 202 wherein the new sequence to be compared is stored to a memory in a computer system 100. As discussed above, the memory could be any type of memory, including RAM or an internal storage device.

The process 200 then moves to a state 204 wherein a database of sequences is opened for analysis and comparison. The process 200 then moves to a state 206 wherein the first sequence stored in the database is read into a memory on the computer. A comparison is then performed at a state 210 to determine if the first sequence is the same as the second sequence. It is important to note that this step is not limited to performing an exact comparison between the new sequence and the first sequence in the database. Well-known methods are known to those of skill in the art for comparing two nucleotide or protein sequences, even if they are not identical. For example, gaps can be introduced into one sequence in order to raise the homology level between the two tested sequences. The parameters that control whether gaps or other features are introduced into a sequence during comparison are normally entered by the user of the computer system.

Once a comparison of the two sequences has been performed at the state 210, a determination is made at a decision state 210 whether the two sequences are the same. Of course, the term "same" is not limited to sequences that are absolutely identical. Sequences that are within the homology parameters entered by the user will be marked as "same" in the process 200.

If a determination is made that the two sequences are the same, the process 200 moves to a state 214 wherein the name of the sequence from the database is displayed to the user. This state notifies the user that the sequence with the displayed name fulfills the homology constraints that were entered. Once the name of the stored sequence is displayed to the user, the process 200 moves to a decision state 218 wherein a determination is made whether more sequences exist in the database. If no more sequences exist in the database, then the process 200 terminates at an end state 220. However, if more sequences do exist in the database, then the process 200 moves to a state 224 wherein a pointer is moved to the next sequence in the database so that it can be compared to the new sequence. In this manner, the new sequence is aligned and compared with every sequence in the database.

It should be noted that if a determination had been made at the decision state 212 that the sequences were not homologous, then the process 200 would move immediately to the decision state 218 in order to determine if any other sequences were available in the database for comparison.

Accordingly, one aspect of the present invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid code of the invention or a polypeptide code of the invention,. In some embodiments the computer system further comprises a data storage device having retrievably stored thereon reference nucleotide sequences or polypeptide sequences to be compared to the nucleic acid code of the invention or polypeptide code of the invention and a sequence comparer for conducting the comparison. For example, the sequence comparer may comprise a computer program which indicates polymorphisms. In other aspects of the computer system, the system further comprises an identifier which identifies features in said sequence. The sequence comparer may indicate a homology level between the sequences compared

or identify motifs implicated in biological function and structural motifs in the nucleic acid code of the invention and polypeptide codes of the invention or it may identify structural motifs in sequences which are compared to these nucleic acid codes and polypeptide codes. In some embodiments, the data storage device may have stored thereon the sequences of at least 2, 5, 10, 15, 20, 25, 30, 50, 75, 100, 150 or 200 of the nucleic acid codes of the invention or polypeptide codes of the invention.

Another aspect of the present invention is a method for determining the level of homology between a nucleic acid code of the invention and a reference nucleotide sequence, comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through the use of a computer program which determines homology levels and determining homology between the nucleic acid code and the reference nucleotide sequence with the computer program. The computer program may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, including BLAST2N with the default parameters or with any modified parameters. The method may be implemented using the computer systems described above. The method may also be performed by reading 2, 5, 10, 15, 20, 25, 30, 50, 75, 100, 150 or 200 of the above described nucleic acid codes of the invention through the use of the computer program and determining homology between the nucleic acid codes and reference nucleotide sequences.

Figure 4 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous. The process 250 begins at a start state 252 and then moves to a state 254 wherein a first sequence to be compared is stored to a memory. The second sequence to be compared is then stored to a memory at a state 256. The process 250 then moves to a state 260 wherein the first character in the first sequence is read and then to a state 262 wherein the first character of the second sequence is read. It should be understood that if the sequence is a nucleotide sequence, then the character would normally be either A, T, C, G or U. If the sequence is a protein sequence, then it should be in the single letter amino acid code so that the first and sequence sequences can be easily compared.

A determination is then made at a decision state 264 whether the two characters are the same. If they are the same, then the process 250 moves to a state 268 wherein the next characters in the first and second sequences are read. A determination is then made whether the next characters are the same. If they are, then the process 250 continues this loop until two characters are not the same. If a determination is made that the next two characters are not the same, the process 250 moves to a decision state 274 to determine whether there are any more characters either sequence to read.

If there aren't any more characters to read, then the process 250 moves to a state 276 wherein the level of homology between the first and second sequences is displayed to the user. The level of homology is determined by calculating the proportion of characters between the sequences

that were the same out of the total number of sequences in the first sequence. Thus, if every character in a first 100 nucleotide sequence aligned with a every character in a second sequence, the homology level would be 100%.

Alternatively, the computer program may be a computer program which compares the nucleotide sequences of the nucleic acid codes of the present invention, to reference nucleotide sequences in order to determine whether the nucleic acid code of the invention differs from a reference nucleic acid sequence at one or more positions. Optionally such a program records the length and identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or the nucleic acid code of the invention. In one embodiment, the computer program may be a program which determines whether the nucleotide sequences of the nucleic acid codes of the invention contain one or more single nucleotide polymorphisms (SNP) with respect to a reference nucleotide sequence. These single nucleotide polymorphisms may each comprise a single base substitution, insertion, or deletion.

Another embodiment of the present invention is a method for comparing a first sequence to a reference sequence wherein the first sequence is selected from the group consisting of a cDNA code of SEQID NOs. 1-297 and a polypeptide code of SEQ ID NOs. 298-594 comprising the steps of reading the first sequence and the reference sequence through use of a computer program which compares sequences and determining differences between the first sequence and the reference sequence with the computer program. In some aspects of this embodiment, said step of determining differences between the first sequence and the reference sequence comprises identifying polymorphisms.

Another aspect of the present invention is a method for determining the level of homology between a polypeptide code of the invention and a reference polypeptide sequence, comprising the steps of reading the polypeptide code of the invention and the reference polypeptide sequence through use of a computer program which determines homology levels and determining homology between the polypeptide code and the reference polypeptide sequence using the computer program.

Accordingly, another aspect of the present invention is a method for determining whether a nucleic acid code of the invention differs at one or more nucleotides from a reference nucleotide sequence comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide sequence with the computer program. In some embodiments, the computer program is a program which identifies single nucleotide polymorphisms The method may be implemented by the computer systems described above and the method illustrated in Figure 4. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30, 50, 75, 100, 150 or 200 of the nucleic acid codes of the invention and the reference nucleotide sequences through the use of the computer

program and identifying differences between the nucleic acid codes and the reference nucleotide sequences with the computer program.

Thus, another embodiment of the present invention is a method for comparing a first sequence to a reference sequence wherein the first sequence is selected from the group consisting of the nucleic acid codes of the present invention or the polypeptide codes of the present invention comprising the steps of reading the first sequence and the reference sequence through use of a computer program which compares sequences and determining differences between the first sequence and the reference sequence with the computer program. In some aspects of this embodiment, said step of determining differences between the first sequence and the reference sequence comprises identifying polymorphisms.

Another aspect of the present invention is a method for determining the level of identity between a first sequence and a reference sequence, wherein the first sequence is selected from the group consisting of the nucleic acid codes of the present invention or the polypeptide codes of the present invention, comprising the steps of reading the first sequence and the reference sequence through the use of a computer program which determines identity levels and determining identity between the first sequence and the reference sequence with the computer program.

In other embodiments the computer based system may further comprise an identifier for identifying features within the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. An "identifier" refers to one or more programs which identifies certain features within the above-described nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. In one embodiment, the identifier may comprise a program which identifies an open reading frame in the cDNAs codes of the invention.

Another embodiment of the present invention is a method for identifying a feature in a sequence selected from the group consisting of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention comprising the steps of reading the sequence through the use of a computer program which identifies features in sequences and identifying features in the sequence with said computer program. In one aspect of this embodiment, the computer program comprises a computer program which identifies open reading frames. In a further embodiment, the computer program comprises a program that identifies linear or structural motifs in a polypeptide sequence.

Figure 5 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence. The process 300 begins at a start state 302 and then moves to a state 304 wherein a first sequence that is to be checked for features is stored to a memory 115 in the computer system 100. The process 300 then moves to a state 306 wherein a database of sequence features is opened. Such a database would include a list of each feature's attributes along with the name of the feature. For example, a feature name could be "Initiation

Codon" and the attribute would be "ATG". Another example would be the feature name "TAATAA Box" and the feature attribute would be "TAATAA". An example of such a database is produced by the University of Wisconsin Genetics Computer Group (www.gcg.com).

Once the database of features is opened at the state 306, the process 300 moves to a state 308 wherein the first feature is read from the database. A comparison of the attribute of the first feature with the first sequence is then made at a state 310. A determination is then made at a decision state 316 whether the attribute of the feature was found in the first sequence. If the attribute was found, then the process 300 moves to a state 318 wherein the name of the found feature is displayed to the user.

The process 300 then moves to a decision state 320 wherein a determination is made whether move features exist in the database. If no more features do exist, then the process 300 terminates at an end state 324. However, if more features do exist in the database, then the process 300 reads the next sequence feature at a state 326 and loops back to the state 310 wherein the attribute of the next feature is compared against the first sequence.

It should be noted, that if the feature attribute is not found in the first sequence at the decision state 316, the process 300 moves directly to the decision state 320 in order to determine if any more features exist in the database.

In another embodiment, the identifier may comprise a molecular modeling program which determines the 3-dimensional structure of the polypeptides codes of the invention. Such programs may use any methods known to those skilled in the art including methods based on homology-modeling, fold recognition and *ab initio* methods as described in Sternberg *et al.*, 1999, which disclosure is hereby incorporated by reference in its entirety. In some embodiments, the molecular modeling program identifies target sequences that are most compatible with profiles representing the structural environments of the residues in known three-dimensional protein structures. (See, e.g., Eisenberg *et al.*, U.S. Patent No. 5,436,850 issued July 25, 1995, which disclosure is hereby incorporated by reference in its entirety). In another technique, the known three-dimensional structures of proteins in a given family are superimposed to define the structurally conserved regions in that family. This protein modeling technique also uses the known three-dimensional structure of a homologous protein to approximate the structure of the polypeptide codes of the invention. (See e.g., Srinivasan, *et al.*, U.S. Patent No. 5,557,535 issued September 17, 1996, which disclosure is hereby incorporated by reference in its entirety). Conventional homology modeling techniques have been used routinely to build models of proteases and antibodies. (Sowdhamini *et al.*, (1997)). Comparative approaches can also be used to develop three-dimensional protein models when the protein of interest has poor sequence identity to template proteins. In some cases, proteins fold into similar three-dimensional structures despite having very weak sequence identities. For example, the three-dimensional structures of a number of helical cytokines fold in similar three-dimensional topology in spite of weak sequence homology.

The recent development of threading methods now enables the identification of likely folding patterns in a number of situations where the structural relatedness between target and template(s) is not detectable at the sequence level. Hybrid methods, in which fold recognition is performed using Multiple Sequence Threading (MST), structural equivalencies are deduced from the threading output using a distance geometry program DRAGON to construct a low resolution model, and a full-atom representation is constructed using a molecular modeling package such as QUANTA.

According to this 3-step approach, candidate templates are first identified by using the novel fold recognition algorithm MST, which is capable of performing simultaneous threading of multiple aligned sequences onto one or more 3-D structures. In a second step, the structural equivalencies obtained from the MST output are converted into interresidue distance restraints and fed into the distance geometry program DRAGON, together with auxiliary information obtained from secondary structure predictions. The program combines the restraints in an unbiased manner and rapidly generates a large number of low resolution model confirmations. In a third step, these low resolution model confirmations are converted into full-atom models and subjected to energy minimization using the molecular modeling package QUANTA. (See e.g., Aszódi *et al.*, (1997)).

The results of the molecular modeling analysis may then be used in rational drug design techniques to identify agents which modulate the activity of the polypeptide codes of the invention.

Accordingly, another aspect of the present invention is a method of identifying a feature within the nucleic acid codes of the invention or the polypeptide codes of the invention comprising reading the nucleic acid code(s) or the polypeptide code(s) through the use of a computer program which identifies features therein and identifying features within the nucleic acid code(s) or polypeptide code(s) with the computer program. In one embodiment, computer program comprises a computer program which identifies open reading frames. In a further embodiment, the computer program identifies linear or structural motifs in a polypeptide sequence. In another embodiment, the computer program comprises a molecular modeling program. The method may be performed by reading a single sequence or at least 2, 5, 10, 15, 20, 25, 30, 50, 75, 100, 150 or 200 of the nucleic acid codes of the invention or the polypeptide codes of the invention through the use of the computer program and identifying features within the nucleic acid codes or polypeptide codes with the computer program.

The nucleic acid codes of the invention or the polypeptide codes of the invention may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, they may be stored as text in a word processing file, such as MicrosoftWORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE. In addition, many computer programs and databases may be used as sequence comparers, identifiers, or sources of reference nucleotide or polypeptide sequences to be compared to the nucleic acid codes of the invention or the polypeptide codes of the

invention. The following list is intended not to limit the invention but to provide guidance to programs and databases which are useful with the nucleic acid codes of the invention or the polypeptide codes of the invention. The programs and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul *et al*, 1990), FASTA (Pearson and Lipman, 1988), FASTDB (Brutlag *et al.*, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius2.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMm (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the EMBL/Swissprotein database, the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwents's World Drug Index database, the BioByteMasterFile database, the Genbank database, and the Genseqn database. Many other programs and data bases would be apparent to one of skill in the art given the present disclosure.

Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

CONCLUSION

As discussed above, the GENSET polynucleotides and polypeptides of the present invention or fragments thereof can be used for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; as a reagent (including a labeled reagent) in assays designed to quantitatively determine levels of GENSET expression in biological samples; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; for selecting and making oligomers for attachment to a

"gene chip" or other support, including for examination for expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris *et al.*, (1993) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The proteins or polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors of the binding interaction. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning; A Laboratory Manual", 2d ed., Cole Spring Harbor Laboratory Press, Sambrook, J., E.F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology; Guide to Molecular Cloning Techniques", Academic Press, Berger and Kimmel eds., 1987, which disclosures are hereby incorporated by reference in their entireties.

Polynucleotides and proteins of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art in view of the disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be defined only by reference to the appended claims.

EXAMPLES

Preparation of Antibody Compositions to GENSET proteins

Substantially pure protein or polypeptide is isolated from transfected or transformed cells containing an expression vector encoding a GENSET protein or a portion thereof. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes in the GENSET protein or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler and Milstein, (1975) or derivative methods thereof. Also see Harlow and Lane. (1988)..

Briefly, a mouse is repetitively inoculated with a few micrograms of the GENSET protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, (1980), which disclosure is hereby incorporated by reference in its entirety, and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, *et al.* (1986) Section 21-2.

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the GENSET protein or a portion thereof can be prepared by immunizing suitable non-human animal with the GENSET protein or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable non-human animal is preferably a non-human mammal is selected, usually a mouse, rat, rabbit, goat, or horse. Alternatively, a crude preparation which has been enriched for GENSET concentration can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e.g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH).

Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art. An effective immunization protocol for rabbits can be found in Vaitukaitis *et al.* (1971), which disclosure is hereby incorporated by reference in its entirety.

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony *et al.*, (1973), which disclosure is hereby incorporated by reference in its entirety. Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 uM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher (1980), which disclosure is hereby incorporated by reference in its entirety.

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

Biological assays

Assaying GENSET Secreted Proteins to Determine Whether they Bind to the Cell Surface

The secreted proteins encoded by the GENSET cDNAs, preferably the proteins of SEQ ID NOs: 242-272 and 274-384, or fragments thereof are cloned into expression vectors. The proteins are purified by size, charge, immunochromatography or other techniques familiar to those skilled in the art. Following purification, the proteins are labeled using techniques known to those skilled in the art. The labeled proteins are incubated with cells or cell lines derived from a variety of organs or tissues to allow the proteins to bind to any receptor present on the cell surface. Following the incubation, the cells are washed to remove non-specifically bound protein. The labeled proteins are detected by autoradiography. Alternatively, unlabeled proteins may be incubated with the cells and detected with antibodies having a detectable label, such as a fluorescent molecule, attached thereto.

Specificity of cell surface binding may be analyzed by conducting a competition analysis in which various amounts of unlabeled protein are incubated along with the labeled protein. The

amount of labeled protein bound to the cell surface decreases as the amount of competitive unlabeled protein increases. As a control, various amounts of an unlabeled protein unrelated to the labeled protein is included in some binding reactions. The amount of labeled protein bound to the cell surface does not decrease in binding reactions containing increasing amounts of unrelated unlabeled protein, indicating that the protein encoded by the cDNA binds specifically to the cell surface.

As discussed herein, secreted proteins have been shown to have a number of important physiological effects and, consequently, represent a valuable therapeutic resource. The secreted proteins encoded by the cDNAs or fragments thereof made using any of the methods described therein may be evaluated to determine their physiological activities as described below.

Assaying GENSET proteins or Fragments Thereof for Cytokine, Cell Proliferation or Cell Differentiation Activity

Secreted proteins may act as cytokines or may affect cellular proliferation or differentiation. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7c and CMK. The proteins encoded by the cDNAs of the invention or fragments thereof may be evaluated for their ability to regulate T cell or thymocyte proliferation in assays such as those described above or in the following references, which are incorporated herein by reference: Current Protocols in Immunology, Ed. by J.E. Coligan *et al.*, Greene Publishing Associates and Wiley-Interscience; Takai *et al. J. Immunol.* 137:3494-3500, 1986. Bertagnolli *et al. J. Immunol.* 145:1706-1712, 1990. Bertagnolli *et al.*, *Cellular Immunology* 133:327-341, 1991. Bertagnolli, *et al. J. Immunol.* 149:3778-3783, 1992; Bowman *et al.*, *J. Immunol.* 152:1756-1761, 1994.

In addition, numerous assays for cytokine production and/or the proliferation of spleen cells, lymph node cells and thymocytes are known. These include the techniques disclosed in Current Protocols in Immunology. J.E. Coligan *et al.* Eds., Vol 1 pp. 3.12.1-3.12.14 John Wiley and Sons, Toronto. 1994; and Schreiber, R.D. Current Protocols in Immunology., *supra* Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

The proteins encoded by the cDNAs of the invention may also be assayed for the ability to regulate the proliferation and differentiation of hematopoietic or lymphopoietic cells. Many assays for such activity are familiar to those skilled in the art, including the assays in the following references, which are incorporated herein by reference: Bottomly, K., Davis, L.S. and Lipsky, P.E., Measurement of Human and Murine Interleukin 2 and Interleukin 4, Current Protocols in

Immunology., J.E. Coligan *et al.* Eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries *et al.*, *J. Exp. Med.* 173:1205-1211, 1991; Moreau *et al.*, *Nature* 36:690-692, 1988; Greenberger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 80:2931-2938, 1983; Nordan, R., Measurement of Mouse and Human Interleukin 6 *Current Protocols in Immunology*. J.E. Coligan *et al.* Eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 83:1857-1861, 1986; Bennett, F., Giannotti, J., Clark, S.C. and Turner, K.J., Measurement of Human Interleukin 11 Current Protocols in Immunology. J.E. Coligan *et al.* Eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Ciarletta, A., Giannotti, J., Clark, S.C. and Turner, K.J., Measurement of Mouse and Human Interleukin 9 Current Protocols in Immunology. J.E. Coligan *et al.*, Eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

The proteins encoded by the cDNAs of the invention may also be assayed for their ability to regulate T-cell responses to antigens. Many assays for such activity are familiar to those skilled in the art, including the assays described in the following references, which are incorporated herein by reference: Chapter 3 (*In vitro* Assays for Mouse Lymphocyte Function), Chapter 6 (Cytokines and Their Cellular Receptors) and Chapter 7, (Immunologic Studies in Humans) in Current Protocols in Immunology, J.E. Coligan *et al.* Eds. Greene Publishing Associates and Wiley-Interscience; Weinberger *et al.*, *Proc. Natl. Acad. Sci. USA* 77:6091-6095, 1980; Weinberger *et al.*, *Eur. J. Immun.* 11:405-411, 1981; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988.

Those proteins which exhibit cytokine, cell proliferation, or cell differentiation activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which induction of cell proliferation or differentiation is beneficial. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

Assaying GENSET proteins or Fragments Thereof for Activity as Immune System Regulators

The proteins encoded by the cDNAs of the invention may also be evaluated for their effects as immune regulators. For example, the proteins may be evaluated for their activity to influence thymocyte or splenocyte cytotoxicity. Numerous assays for such activity are familiar to those skilled in the art including the assays described in the following references, which are incorporated herein by reference: Chapter 3 (*In vitro* Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic studies in Humans) in Current Protocols in Immunology, J.E. Coligan *et al.* Eds, Greene Publishing Associates and Wiley-Interscience; Herrmann *et al.*, *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981; Herrmann *et al.*, *J. Immunol.* 128:1968-1974, 1982; Handa *et al.*, *J. Immunol.* 135:1564-1572, 1985; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988; Herrmann *et al.*, *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981;

Herrmann *et al.*, *J. Immunol.* 128:1968-1974, 1982; Handa *et al.*, *J. Immunol.* 135:1564-1572, 1985; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Bowman *et al.*, *J. Virology* 61:1992-1998; Takai *et al.*, *J. Immunol.* 140:508-512, 1988; Bertagnolli *et al.*, *Cellular Immunology* 133:327-341, 1991; Brown *et al.*, *J. Immunol.* 153:3079-3092, 1994.

The proteins encoded by the cDNAs of the invention may also be evaluated for their effects on T-cell dependent immunoglobulin responses and isotype switching. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Maliszewski, *J. Immunol.* 144:3028-3033, 1990; Mond, J.J. and Brunswick, M Assays for B Cell Function: *In vitro* Antibody Production, Vol 1 pp. 3.8.1-3.8.16 in Current Protocols in Immunology. J.E. Coligan *et al* Eds., John Wiley and Sons, Toronto. 1994.

The proteins encoded by the cDNAs of the invention may also be evaluated for their effect on immune effector cells, including their effect on Th1 cells and cytotoxic lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Chapter 3 (*In vitro* Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic Studies in Humans) in Current Protocols in Immunology, J.E. Coligan *et al.* Eds., Greene Publishing Associates and Wiley-Interscience; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*; *J. Immunol.* 140:508-512, 1988; Bertagnolli *et al.*, *J. Immunol.* 149:3778-3783, 1992.

The proteins encoded by the cDNAs of the invention may also be evaluated for their effect on dendritic cell mediated activation of naive T-cells. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Guery *et al.*, *J. Immunol.* 134:536-544, 1995; Inaba *et al.*, *Journal of Experimental Medicine* 173:549-559, 1991; Macatonia *et al.*, *Journal of Immunology* 154:5071-5079, 1995; Porgador *et al.*, *Journal of Experimental Medicine* 182:255-260, 1995; Nair *et al.*, *Journal of Virology* 67:4062-4069, 1993; Huang *et al.*, *Science* 264:961-965, 1994; Macatonia *et al.*, *Journal of Experimental Medicine* 169:1255-1264, 1989; Bhardwaj *et al.*, *Journal of Clinical Investigation* 94:797-807, 1994; and Inaba *et al.*, *Journal of Experimental Medicine* 172:631-640, 1990.

The proteins encoded by the cDNAs of the invention may also be evaluated for their influence on the lifetime of lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Darzynkiewicz *et al.*, *Cytometry* 13:795-808, 1992; Gorczyca *et al.*, *Leukemia* 7:659-670, 1993; Gorczyca *et al.*, *Cancer Research* 53:1945-1951, 1993; Itoh *et al.*, *Cell* 66:233-243, 1991; Zacharchuk, *Journal of Immunology* 145:4037-4045, 1990; Zamai *et al.*, Cytometry 14:891-897, 1993; Gorczyca *et al.*, *International Journal of Oncology* 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica *et al.*, Blood 84:111-117, 1994; Fine *et al.*, Cellular immunology 155:111-122, 1994; Galy *et al.*, Blood 85:2770-2778, 1995; Toki *et al.*, Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

Those proteins which exhibit activity as immune system regulators activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of immune activity is beneficial. For example, the protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to regulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T-cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al.*, Science 257:789-792 (1992) and Turka *et al.*, Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a

number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/pr/pr mice or NZB hybrid mice, murine autoimmuno collagen arthritis, diabetes mellitus in OD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory form of B lymphocyte antigens systemically.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the *in vitro* activated T cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to T cells *in vivo*, thereby activating the T cells.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules, can be transfected with nucleic acids encoding all or a fragment of (e.g., a cytoplasmic-domain truncated fragment) of an MHC class I α chain protein and β_2 microglobulin protein or an MHC class II α chain protein and an MHC class II β chain protein to thereby express MHC class I or MHC class II proteins on the cell surface.

Expression of the appropriate class II or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain,can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

Assaying GENSET proteins or Fragments Thereof for Hematopoiesis Regulating Activity

The proteins encoded by the cDNAs of the invention or fragments thereof may also be evaluated for their hematopoiesis regulating activity. For example, the effect of the proteins on embryonic stem cell differentiation may be evaluated. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Johansson *et al. Cellular Biology* 15:141-151, 1995; Keller *et al., Molecular and Cellular Biology* 13:473-486, 1993; McClanahan *et al., Blood* 81:2903-2915, 1993.

The proteins encoded by the cDNAs of the invention or fragments thereof may also be evaluated for their influence on the lifetime of stem cells and stem cell differentiation. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Freshney, M.G. Methylcellulose Colony Forming Assays, in Culture of Hematopoietic Cells. R.I. Freshney, *et al.* Eds. pp. 265-268, Wiley-Liss, Inc., New York, NY. 1994; Hirayama *et al., Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; McNiece, I.K. and Briddell, R.A. Primitive Hematopoietic Colony Forming Cells with High Proliferative Potential, in Culture of Hematopoietic Cells. R.I. Freshney, *et al.* eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, NY. 1994; Neben *et al., Experimental Hematology* 22:353-359, 1994; Ploemacher, R.E. Cobblestone Area Forming Cell Assay, In Culture of Hematopoietic Cells. R.I. Freshney, *et al.* Eds. pp. 1-21, Wiley-Liss, Inc., New York, NY. 1994; Spooncer, E., Dexter, M. and Allen, T. Long Term Bone Marrow Cultures in the Presence of Stromal Cells, in Culture of Hematopoietic Cells. R.I. Freshney, *et al.* Eds. pp. 163-179, Wiley-Liss, Inc., New York, NY. 1994; and Sutherland, H.J. Long Term Culture Initiating Cell Assay, in Culture of Hematopoietic Cells. R.I. Freshney, *et al.* Eds. pp. 139-162, Wiley-Liss, Inc., New York, NY. 1994.

Those proteins which exhibit hematopoiesis regulatory activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of hematopoeisis is

beneficial. For example, a protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantion, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

Assaying GENSET proteins or Fragments Thereof for Regulation of Tissue Growth

The proteins encoded by the cDNAs of the invention or fragments thereof may also be evaluated for their effect on tissue growth. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in International Patent Publication No. WO95/16035, International Patent Publication No. WO95/05846 and International Patent Publication No. WO91/07491, which are incorporated herein by reference.

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H1 and Rovee, DT, eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978) which are incorporated herein by reference.

Those proteins which are involved in the regulation of tissue growth may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of tissue growth is beneficial. For example, a protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration,

as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e., for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve

degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium) muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to generate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokinc damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

Assaying GENSET proteins or Fragments Thereof for Regulation of Reproductive Hormones or Cell Movement

The proteins encoded by the cDNAs of the invention or fragments thereof may also be evaluated for their ability to regulate reproductive hormones, such as follicle stimulating hormone. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Vale *et al.*, *Endocrinology* 91:562-572, 1972; Ling *et al.*, *Nature* 321:779-782, 1986; Vale *et al.*, *Nature* 321:776-779, 1986; Mason *et al.*, *Nature* 318:659-663, 1985; Forage *et al.*, *Proc. Natl. Acad. Sci. USA* 83:3091-3095, 1986. Chapter 6.12 (Measurement of Alpha and Beta Chemokines) Current Protocols in Immunology, J.E. Coligan *et al.* Eds. Greene Publishing Associates and Wiley-

Interscience ; Taub *et al. J. Clin. Invest.* 95:1370-1376, 1995; Lind *et al. APMIS* 103:140-146, 1995; Muller *et al. Eur. J. Immunol.* 25:1744-1748; Gruber *et al. J. of Immunol.* 152:5860-5867, 1994; Johnston *et al. J. of Immunol.* 153:1762-1768, 1994.

Those proteins which exhibit activity as reproductive hormones or regulators of cell movement may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of reproductive hormones or cell movement are beneficial. For example, a protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of folic stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin α family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-B group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, United States Patent 4,798,885, the disclosure of which is incorporated herein by reference. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

Assaying GENSET proteins or Fragments Thereof for Chemotactic/Chemokinetic Activity

The proteins encoded by the cDNAs of the invention or fragments thereof may also be evaluated for chemotactic/chemokinetic activity. For example, a protein of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, cosinophils, epithelial and/or endothelial cells. Chemotactic and chmokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells.

Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhension of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokincs 6.12.1-6.12.28; Taub *et al.* J. Clin. Invest. 95:1370-1376, 1995; Lind *et al.* APMIS 103:140-146, 1995; Mueller *et al* Eur. J. Immunol. 25:1744-1748; Gruber *et al.* J. of Immunol. 152:5860-5867, 1994; Johnston *et al.* J. of Immunol, 153:1762-1768, 1994.

Assaying GENSET proteins or Fragments Thereof for Regulation of Blood Clotting

The proteins encoded by the cDNAs of the invention or fragments thereof may also be evaluated for their effects on blood clotting. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Linet *et al.*, *J. Clin. Pharmacol.* 26:131-140, 1986; Burdick *et al.*, *Thrombosis Res.* 45:413-419, 1987; Humphrey *et al.*, *Fibrinolysis* 5:71-79 (1991); Schaub, *Prostaglandins* 35:467-474, 1988.

Those proteins which are involved in the regulation of blood clotting may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of blood clotting is beneficial. For example, a protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulations disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke)). Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

Assaying GENSET proteins or Fragments Thereof for Involvement in Receptor/Ligand Interactions

The proteins encoded by the cDNAs or a fragment thereof may also be evaluated for their involvement in receptor/ligand interactions. Numerous assays for such involvement are familiar to

those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Chapter 7.28 (Measurement of Cellular Adhesion under Static Conditions 7.28.1-7.28.22) in Current Protocols in Immunology, J.E. Coligan *et al.* Eds. Greene Publishing Associates and Wiley-Interscience; Takai *et al.*, *Proc. Natl. Acad. Sci. USA* 84:6864-6868, 1987; Bierer *et al.*, *J. Exp. Med.* 168:1145-1156, 1988; Rosenstein *et al.*, *J. Exp. Med.* 169:149-160, 1989; Stoltenborg *et al.*, *J. Immunol. Methods* 175:59-68, 1994; Stitt *et al.*, *Cell* 80:661-670, 1995; Gyuris *et al.*, *Cell* 75:791-803, 1993.

For example, the proteins of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune respones). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

Assaying GENSET proteins or Fragments Thereof for Anti-Inflammatory Activity

The proteins encoded by the cDNAs or a fragment thereof may also be evaluated for anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusioninury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

Assaying GENSET proteins or Fragments Thereof for Tumor Inhibition Activity

The proteins encoded by the cDNAs of the invention or a fragment thereof may also be evaluated for tumor inhibition activity. In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-

tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, climinating or inhibiting factors, agents or cell types which promote tumor growth.

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or climination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

References

Abbondanzo *et al.*, (1993), *Meth. Enzymol.*, Academic Press, New York, pp 803-823

Altschul *et al.*, (1990), *J. Mol. Biol.* 215(3):403-410

Altschul *et al.*, (1993), *Nature Genetics* 3:266-272

Altschul *et al.*, (1997), *Nuc. Acids Res.* 25:3389-3402

Ames *et al.*, (1995), *J. Immunol. Meth.* 184:177-186.

Anton and Graham, (1995), *J. Virol.*, 69: 4600-4606

Araki *et al.*, (1995) Proc. Natl. Acad. Sci. U S A. 92(1):160-4.

Ashkenazi *et al.*, (1991), Proc. Natl. Acad. Sci. U S A 88:10535-10539.

Aszódi *et al.*, (1997) Proteins: Structure, Function, and Genetics, Supplement 1:38-42

Attwood *et al.*, (1996) Nucleic Acids Res. 24(1):182-8.

Attwood *et al.*, (2000) Nucleic Acids Res. 28(1):225-7

Bartunek *et al.*, (1996), *Cytokine.* 8(1):14-20.

Bateman *et al.*, (2000) Nucleic Acids Res. 28(1):263-6

Baubonis (1993) *Nucleic Acids Res.* 21(9):2025-9.

Beaucage *et al.*, (1981) *Tetrahedron Lett*, 22: 1859-1862

Benham *et al.* (1989) *Genomics* 4:509-517,

Better *et al.*, (1988), *Science.* 240:1041-1043.

Bhagwat *et al.* (1999) *J. Biol. Chem.* 274:24014-22

Bittle *et al.*, (1985), *Virol.* 66:2347-2354.

Bowie *et al.*, (1994), Science. 247:1306-1310.

Bradley (1987), Production and analysis of chimaeric mice. *In*: E.J. Robertson (Ed.), Teratocarcinomas and embryonic stem cells: A practical approach. IRL Press, Oxford, pp.113.

Bram *et al.*, (1993), *Mol. Cell Biol.,* 13 : 4760-4769

Brinkman *et al.*, (1995*) J. Immunol Methods.* 182:41-50.

Brown *et al.*, (1979) *Meth. Enzymol.* 68:109-151

Brutlag *et al.* (1990) *Comp. App. Biosci.* 6:237-245

Bucher and Bairoch (1994) Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology. Altman *et al*, Eds., pp53-61, AAAIPress, Menlo Park.

Burton *et al.* (1994), *Adv. Immunol.* 57:191-280

Bush *et al.*, (1997), J. Chromatogr., 777 : 311-328.

Butt and Karathanasis (1995*) Gene Expr.* 4(6):319-36.

Carlson *et al.*, (1997*), J. Biol. Chem.* 272(17):11295-11301.

Chai *et al.*, (1993) Biotechnol. Appl. Biochem. 18:259-273.

Chang *et al.*, (1993) *Gene* 127:95-8

Chee *et al.*, (1996) *Science.* 274:610-614.

Chen *et al.* (1987) *Mol. Cell. Biol.* 7:2745-2752.

Chen *et al.*, (1998), *Cancer Res.* 58(16):3668-3678.

Cherif *et al.*, (1990) Proc. Natl. Acad. Sci. U.S.A., 87:6639-6643

Cho *et al.*, (1998), *Proc. Natl. Acad. Sci. USA*, 95(7) : 3752-3757.

Chou, (1989), *Mol. Endocrinol.* 3: 1511-1514.

Chow *et al.*, (1985), Proc. Natl. Acad. Sci. USA. 82:910-914.

Christophe *et al.* (2000) *Cell Signal* 12(5):337-41

Ciminale *et al.* (1999) *Oncogene* 18:4505-14

Cleland *et al.*, (1993), Crit. Rev. Therapeutic Drug Carrier Systems. 10:307-377.

Coles *et al.*, (1998) *Hum Mol Genet* 7:791-800

Compton (1991) *Nature.* 350(6313):91-92.

Corpet *et al.* (2000) Nucleic Acids Res. 28(1):267-9

Cox *et al.*, (1990) *Science* 250:245-250

Creighton (1983), Proteins: Structures and Molecular Principles, W.H. Freeman & Co. 2nd Ed., T. E., New York

Creighton, (1993) , Posttranslational Covalent Modification of Proteins, W.H. Freeman and Company, New York B.C. Johnson, Ed., Academic Press, New York 1-12

Cunningham *et al.* (1989), Science 244:1081-1085.

Davis *et al.*, (1986) Basic Methods in Molecular Biology, ed., Elsevier Press, NY,

Decker and Parker, (1995) Curr. Opin. Cell. Biol. 7(3) :368-92

Dempster *et al.*, (1977) Stat. Soc., 39B:1-38.

Deng *et al.*, (1998) Blood. 92(6):1981-1988.

Dent and Latchman (1993) The DNA mobility shift assay. In: Transcription Factors: A Practical Approach (Latchman DS, ed.) pp1-26. Oxford: IRL Press

Derrigo *et al.*, (2000) Int. J. Mol. Med. 5(2) :111-23

Eckner *et al.*, (1991) EMBO J. 10:3513-3522.

Edwards and Leatherbarrow, (1997) Analytical Biochemistry, 246, 1-6

Engvall, (1980) Meth. Enzymol. 70:419

Erlich, (1992) PCR Technology; Principles and Applications for DNA Amplification. W.H. Freeman and Co., New York

Essl *et al.,* (1999) *FEBS Lett.* 453:169-73.

Feldman and Steg, (1996), Medecine/Sciences, 12:47-55

Felgner (1996) Hum Gene Ther. 7(15):1791-3.

Felici, (1991), J. Mol. Biol., 222:301-310

Fell *et al.*, (1991), J. Immunol. 146:2446-2452.

Fields and Song, (1989), Nature, 340: 245-246

Fisher, (1980) Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C.

Flotte *et al.*, (1992) Am. J. Respir. Cell Mol. Biol. 7:349-356.

Fodor *et al.*, (1991) Science 251:767-777.

Foster *et al.*, (1996) Genomics 33:185-192

Fountoulakis *et al.*, (1995) Biochem. 270:3958-3964.

Fraley *et al.*, (1979) Proc. Natl. Acad. Sci. USA. 76:3348-3352.

Frazer *et al.*, (1992) Genomics 14:574-584

Fried and Crothers, (1981) Nucleic Acids Res. 9:6505-6525

Fromont-Racine *et al.*, (1997), Nature Genetics, 16(3) : 277-282.

Fry *et al.*, (1992) Biotechniques, 13: 124-131

Fudenberg, (1980) Chap. 26 in: Basic 503 Clinical Immunology, 3rd Ed. Lange, Los Altos, California

Fuller S. A. *et al.* (1996) Immunology in Current Protocols in Molecular Biology,

Furth P.A. *et al.* (1994) Proc. Natl. Acad. Sci USA. 91:9302-9306.

Garner and Revzin, (1981) Nucleic Acids Res 9:3047-3060

Gentz *et al.*, (1989) Proc Natl Acad Sci U S A. 86(3):821-4.

Geysen *et al.*, (1984), Proc. Natl. Acad. Sci. U.S.A. 81:3998-4002.

Ghosh and Bacchawat, (1991), Targeting of liposomes to hepatocytes, IN: Liver Diseases, Targeted diagnosis and therapy using specific rceptors and ligands. Eds., Marcel Dekeker, New York, pp. 87-104.

Gillies *et al.*, (1989), J. Immunol Methods. 125:191-202.

Gillies *et al.*, (1992), Proc Natl Acad Sci U S A 89:1428-1432.

Glaser *et al.* (1998*) Plant Mol Biol* 38:311-38;

Gonnet *et al.*, (1992), Science 256:1443-1445

Gopal (1985) Mol. Cell. Biol., 5:1188-1190.

Gossen *et al.*, (1992) Proc. Natl. Acad. Sci. USA. 89:5547-5551.

Gossen *et al.*, (1995) Science. 268:1766-1769.

Graham *et al.*, (1973) Virol. 52:456-457.

Green *et al.*, (1986) Ann. Rev. Biochem. 55:569-597

Greenspan and Bona (1989), FASEB J. 7(5):437-444.

Griffais *et al.*, (1991) Nucleic Acids Res. 19: 3887-3891

Griffin *et al.*, (1989) Science 245:967-971

Gu H. *et al.*, (1993) Cell 73:1155-1164.

Gu H. *et al.*, (1994) Science 265:103-106.

Guatelli *et al.*, (1990) Proc. Natl. Acad. Sci. USA. 35:273-286.

Gyuris *et al.*, (1993) Cell 75:791-803

Hames and Higgins (1985) Nucleic Acid Hybridization: A Practical Approach. Hames and Higgins Ed., IRL Press, Oxford.

Hammerling (1981), Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y. 563-681.

Hansson *et al.*, (1999), J. Mol. Biol. 287:265-276.

Haravama (1998), Trends Biotechnol. 16(2): 76-82.

Harland *et al.*, (1985) J. Cell. Biol. 101:1094-1095.

Harlow and Lane, (1988) Antibodies A Laboratory Manual. Cold Spring Harbor Laboratory. pp. 53-242

Harper *et al.*, (1993), Cell, 75 : 805-816

Harrop *et al.*, (1998), J. Immunol. 161(4):1786-1794.

Haynes *et al.*, (1996) J Biotechnol. 44(1-3):37-42.

Henikoff and Henikoff, (1993), Proteins 17:49-61

Henikoff *et al.*, (2000) Electrophoresis 21(9):1700-6

Henikoff *et al.*, (2000) Nucleic Acids Res. 28(1):228-30

Herrman and Neupert, (2000) Curr. Opinion Microbiol. 3:210-4

Higgins *et al.*, (1996), Meth. Enzymol. 266:383-402

Hillier and Green (1991) PCR Methods Appl., 1: 124-8.

Hoess *et al.*, (1986) Nucleic Acids Res. 14:2287-2300.

Hofmann *et al.*, (1999) Nucl. Acids Res. 27:215-219.;

Holm and Sander (1996) Nucleic Acids Res. 24(1):206-9

Holm and Sander (1997) Nucleic Acids Res. 25(1):231-4

Holm and Sander (1999) Nucleic Acids Res. 27(1):244-7

Hoppe *et al.*, (1994), FEBS Letters. 344:191.

Houghten (1985), Proc. Natl. Acad. Sci. USA 82:5131-5135.

Huang *et al.*, (1996) Cancer Res 56(5):1137-1141.

Hunkapiller *et al.*, (1984) Nature. 310(5973):105-11.

Huston *et al.*, (1991), Meth. Enymol. 203:46_88.

Huygen *et al.*, (1996) Nature Medicine. 2(8):893-898.

Izant and Weintraub, (1984) Cell 36(4):1007-15

Jameson and Wolf, (1988), Comp. Appl. Biosci. 4:181-186

Julan *et al.*, (1992) J. Gen. Virol. 73:3251-3255.

Kanegae *et al.*, (1995) Nucl. Acids Res. 23:3816-3821.

Karlin and Altschul, (1990), Proc. Natl. Acad. Sci. USA 87:2267-2268

Kettleborough *et al.*, (1994), Eur. L Immunol. 24:952-958.

Kim U-J. *et al.*, (1996) Genomics 34:213-218.

Klein *et al.*, (1987) Nature. 327:70-73.

Kohler and Milstein, (1975) Nature 256:495

Koller *et al.*; (1992) Annu. Rev. Immunol. 10:705-730.

Kostelny *et al.*, (1992), J. Immunol. 148:1547-1553.

Landschulz *et al.*, (1988), Science. 240:1759.

Ledbetter *et al.*, (1990) Genomics 6:475-481

Lenhard *et al.*, (1996) Gene. 169:187-190.

Levy *et al.*, (1996) Gene Ther. 3(3):201-11.

Lewin, (1989), Proc. Natl. Acad. Sci. USA 86:9832-8935.

Liautard *et al.*, (1997), Cytokine. 9(4):233-241.

Linton *et al.*, (1993) J. Clin. Invest. 92:3029-3037.

Liu *et al.*, (1994) Proc. Natl. Acad. Sci. USA. 91: 4528-4262.

Lo Conte *et al.*, (2000) Nucleic Acids Res. 28(1):257-9.

Lockhart *et al.*, (1996) Nature Biotechnology 14: 1675-1680

Lorenzo and Blasco (1998) Biotechniques. 24(2):308-313.

Lucas (1994), In : Development and Clinical Uses of Haempophilus b Conjugate;

Luetterforst et al., (1999) *J. Cell. Biol.* 145:1443-59

Makrides, (1999) Protein Expr. Purif. 17(2) :183-202

Malik *et al.*, (1992), Exp. Hematol. 20:1028-1035.

Mansour *et al.*, (1988) Nature. 336:348-352.

Marshall *et al.*, (1994) PCR Methods and Applications. 4:80-84.

Maurer *et al.*, (1999) Mol Membr Biol. 16(1):129-40.

McCormick *et al.*, (1994) Genet. Anal. Tech. Appl. 11:158-164.

McLaughlin *et al.*, (1996) Am. J. Hum. Genet. 59:561-569.

Miller and Whelan, (1997) Hum Gene Ther. 8(7):803-15.

Muller *et al.*, (1998), Structure. 6(9):1153-1167.

Mullinax *et al.*, (1992), BioTechniques. 12(6):864-869.

Munro, (1998) *Trends Cell Biol.* 8:11-15

Munro and Pelham (1987) *Cell* 48(5):899-907

Murphy (1997*) Trends Biotechnol.* 15:326-30

Murvai *et al.*, (2000) Nucleic Acids Res. 28(1):260-2

Murzin *et al.*, (1995) J Mol Biol. 247(4):536-40

Muzyczka *et al.*, (1992) Curr. Topics in Micro. and Immunol. 158:97-129.

Nada *et al.*, (1993) Cell 73:1125-1135.

Nagaraja *et al*, (1997. Genome Research 7:210-222

Nagy *et al.*, (1993), Proc. Natl. Acad. Sci. USA 90: 8424-8428.

Nakai and Horton, (1999) Trends Biochem. Sci., 24:34-36

Nakai and Kanehisa (1992) Genomics 14, 897-911

Naramura *et al.*, (1994), Immunol. Lett. 39:91-99.

Narang *et al.*, (1979), Methods Enzymol 68:90-98

Neda *et al.*, (1991) J. Biol. Chem. 266:14143-14146.

Nevill-Manning *et al.*, (1998) Proc. Natl. Acad. Sci. U S A. 95, 5865-5871

Nicolau *et al.*, (1982) Biochim. Biophys. Acta. 721:185-190.

Nicolau *et al.*, (1987), Meth. Enzymol., 149:157-76.

Nissinoff, (1991), J. Immunol. 147(8): 2429-2438.

O'Reilly *et al.*, (1992) Baculovirus Expression Vectors: A Laboratory Manual. W. H. Freeman and Co., New York.

Obermayr *et al.*, (1996) Eur. J. Hum. Genet. 4:242-245

Ohno *et al.*, (1994) Science. 265:781-784.

Oi *et al.*, (1986), BioTechniques 4:214.

Oldenburg *et al.*, (1992), Proc. Natl. Acad. Sci. USA 89:5393-5397.

Orengo *et al.*, (1997) Structure. 5(8):1093-108

Ouchterlony *et al.*, (1973) Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell

Padlan, (1991), Molec. Immunol. 28(4/5):489-498.

Parmley and Smith, (1988) Gene 73:305-318

Patten, *et al.* (1997), Curr Opinion Biotechnol. 8:724-733.

Pearl *et al.*, (2000) Biochem Soc Trans. 28(2):269-75

Pearson and Lipman, (1988), Proc. Natl. Acad. Sci. USA 85(8):2444-2448

Pease and William, (1990), Exp. Cell. Res. 190: 209-211.

Pelham (1990) *Trends Biochem Sci* 12:483-6

Persic *et al.*, (1997), Gene. 1879-81

Pesole *et al.*, (2000) Nucleic Acids Res, 28(1):193-196

Peterson *et al.*, (1993), Proc. Natl. Acad. Sci. USA, 90 : 7593-7597.

Picetti and Borrelli, (2000) *Exp Cell Res* 255:258-69

Pidoux and Armstrong (1992) *EMBO J* Apr;11(4):1583-91

Pietu *et al.*, (1996) Genome Research 6:492-503

Pinckard *et al.*, (1967), Clin. Exp. Immunol 2:331-340.

Pitard *et al.*, (1997), J. Immunol. Methods. 205(2):177-190.

Pongor *et al.* (1993) Protein Eng. 6(4):391-5

Potter *et al.*, (1984) Proc. Natl. Acad. Sci. U.S.A. 81(22):7161-7165.

Prat *et al.*, (1998), J. Cell. Sci. 111(Pt2):237-247.

Raeymaekers *et al.*, (1995) Genomics 29:170-178

Ramunsen *et al.*, (1997), Electrophoresis, 18 : 588-598.

Rattan *et al.*, (1992) Ann NY Acad Sci 663:48-62

Reid *et al.*, (1990) Proc. Natl. Acad. Sci. U.S.A. 87:4299-4303.

Robbins *et al.*, (1987), Diabetes. 36:838-845.

Robertson, (1987), Embryo-derived stem cell lines. In: E.J. Robertson Ed. Teratocarcinomas and embrionic stem cells: a practical approach. IRL Press, Oxford, pp. 71.

Roguska *et al.*, (1994), Proc. Natl. Acad. Sci. U.S.A. 91:969-973.

Ron *et al.*, (1993), Biol Chem., 268 2984-2988.

Rose *et al.*, (1980) Chap. 12 in: Methods in Immunodiagnosis, 2d Ed. John Wiley 503 Sons, New York

Rossi *et al.*, (1991) Pharmacol. Ther. 50:245-254,

Roth *et al.*, (1996) Nature Medicine. 2(9):985-991.

Roux *et al.*, (1989) Proc. Natl. Acad. Sci. U.S.A. 86:9079-9083.

Sambrook *et al.*, (1989) Molecular Cloning: A Laboratory Manual. 2ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Samson *et al.*, (1996) Nature, 382(6593):722-725.

Samulski *et al.*, (1989) J. Virol. 63:3822-3828.

Sanchez-Pescador (1988) J. Clin. Microbiol. 26(10):1934-1938.

Sander and Schneider (1991) Proteins. 9(1):56-68.)

Sauer *et al.*, (1988) Proc. Natl. Acad. Sci. U.S.A. 85:5166-5170.

Sawai *et al.*, (1995), AJRI 34:26-34.

Schedl *et al.*, (1993b), Nucleic Acids Res., 21: 4783-4787.

Schedl *et al.*, 1(993a), Nature, 362: 258-261.

Schena *et al.* (1995) Science 270:467-470

Schena *et al.*, (1996), Proc Natl Acad Sci U S A,.93(20):10614-10619.

Schuler *et al.*, (1996) Science 274:540-546

Schultz *et al.*, (1998) Proc Natl Acad Sci U S A 95, 5857-5864

Schwartz and Dayhoff, (1978), eds., Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundation

Sczakiel *et al.*, (1995) Trends Microbiol. 3(6):213-217.

Seifter *et al.*, (1990) Meth Enzymol 182:626-646

Shay *et al.*, (1991), Biochem. Biophys. Acta, 1072: 1-7.

Shizuya *et al.*, (1992) Proc. Natl. Acad. Sci. U.S.A. 89:8794-8797.

Shu *et al.*, (1993), Proc. Natl. Acad. Sci. U.S.A. 90:7995-7999.

Skerra *et al.*, (1988), Science 240:1038-1040.

Smith and Johnson (1988) Gene. 67(1):31-40.

Smith *et al.*, (1983) Mol. Cell. Biol. 3:2156-2165.

Smith *et al.*, (1996) Antiviral Res. 32(2):99-115.

Sonnhammer and Kahn D (1994) Protein Sci. 3(3):482-92

Sonnhammer *et al.*, (1997) Proteins. 28(3):405-20

Sosnowski, *et al.*, (1997) Proc Natl Acad Sci U S A 94:1119-1123

Sowdhamini *et al.*, (1997) Protein Engineering 10:207, 215

Sternberg (1994) Mamm. Genome. 5:397-404.

Sternberg (1992) Trends Genet. 8:1-16.

Sternberg *et al.*, (1999) Curr Opin Struct Biol. 9(3):368-73.

Stone *et al.*, (2000) J Endocrinol. 164(2):103-18.

Stryer, (1995) Biochemistry, 4th edition

Studnicka *et al.*, (1994), Protein Engineering. 7(6):805-814.

Sutcliffe *et al.*, (1983), Science. 219:660-666.

Szabo *et al.*, (1995) Curr Opin Struct Biol 5, 699-705

Tacson *et al.*, (1996) Nature Medicine. 2(8):888-892.

Taryman *et al.*, (1995), Neuron. 14(4):755-762.

Tatusov *et al.*, (1997) Science, 278, 631 :637

Tatusov *et al.*, (2000) Nucleic Acids Res. 28(1):33-6.)

Te Riele *et al.*, (1990) Nature. 348:649-651.

Thomas *et al.*, (1986) Cell. 44:419-428.

Thomas *et al.*, (1987) Cell. 51:503-512.

Thompson *et al.*, (1994), Nucleic Acids Res. 22(2):4673-4680

Traunecker *et al.*, (1988), Nature. 331:84-86.

Tur-Kaspa *et al.*, (1986) Mol. Cell. Biol. 6:716-718.

Tutt *et al.*, (1991), J. Immunol. 147:60-69.

Ugur and Jones (2000) *Mol Cell Biol* 11:1432-32

Urdea (1988) Nucleic Acids Research. 11:4937-4957.

Urdea *et al.*, (1991) Nucleic Acids Symp. Ser. 24:197-200.

Vaitukaitis *et al.*, (1971) J. Clin. Endocrinol. Metab. 33:988-991

Valadon *et al.*, (1996), J. Mol. Biol., 261:11-22.

Van der Lugt *et al.*, (1991) Gene. 105:263-267.

Vil *et al.*, (1992) Proc Natl Acad Sci U S 89:11337-11341.

Vlasak *et al.*, (1983) Eur. J. Biochem. 135:123-126.

Wabiko *et al.*, (1986) DNA.5(4):305-314.

Wagner *et al.*, (1996) Nat Biotechnol. 14(7):840-4.

Walker *et al.*, (1996) Clin. Chem. 42:9-13.

Wang *et al.*, (1997), Chromatographia, 44 : 205-208.

Ward and Moss, (2000*) J. Virol.* 74:3771-80

Warrington *et al.*, (1991) Genomics 11:701-708

Watson and Pessin, (2000) *J Biol Chem* 275:1261-8

Westerink, (1995), Proc. Natl. Acad. Sci USA., 92:4021-4025

White (1997) B.A. Ed. in Methods in Molecular Biology 67: Humana Press, Totowa

White *et al.* (1997) Genomics. 12:301-306.

Wilson *et al.*, (1984) Cell. 37(3):767-78.

Wong *et al.*, (1980) Gene. 10:87-94.

Wood *et al.*, (1985) Proc. Natl. Acad. Sci. USA 82(6):1585-1588

Wood *et al.*, (1993), Proc. Natl. Acad. Sci. USA, 90: 4582-4585.

Wu and Ataai, (2000) Curr Opin Biotechnol. 11(2):205-8.

Wu and Wu, (1987) J. Biol. Chem. 262:4429-4432.

Wu and Wu, (1988) Biochemistry. 27:887-892.

Yagi T. *et al.*, (1990) Proc. Natl. Acad. Sci. U.S.A. 87:9918-9922.

Yona *et al.*, (1999) Proteins. 37(3):360-78

Yoon *et al*, (1998), J. Immunol. 160(7):3170-3179.

Zhan et al., (1998) *Cancer Immunol Immunother* 46:55-60

Zheng, X.X. *et al.* (1995), J. Immunol. 154:5590-5600.

Zhu *et al.*, (1998), Cancer Res. 58(15):3209-3214.

Zou *et al.*, (1994) Curr. Biol. 4:1099-1103.

Throughout this application, various publications, patents and published patent applications are cited. The disclosures of these publications, patents and published patent specification referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

Table I

Seq Id No	Internal designation	Type	Vector
1	119-003-4-0-C2-CS	DNA	pBluescriptII SK-
2	105-016-1-0-D3-CS	DNA	pBluescriptII SK-
3	105-016-3-0-G10-CS	DNA	pBluescriptII SK-
4	105-026-1-0-A5-CS	DNA	pBluescriptII SK-
5	105-031-1-0-A11-CS	DNA	pBluescriptII SK-
6	105-031-2-0-D3-CS	DNA	pBluescriptII SK-
7	105-035-2-0-C6-CS	DNA	pBluescriptII SK-
8	105-037-2-0-H11-CS	DNA	pBluescriptII SK-
9	105-053-4-0-E8-CS	DNA	pBluescriptII SK-
10	105-074-3-0-H10-CS	DNA	pBluescriptII SK-
11	105-089-3-0-G10-CS	DNA	pBluescriptII SK-
12	105-095-2-0-G11-CS	DNA	pBluescriptII SK-
13	106-006-1-0-E3-CS	DNA	pBluescriptII SK-
14	106-037-1-0-E9-CS.cor	DNA	pBluescriptII SK-
15	106-037-1-0-E9-CS.fr	DNA	pBluescriptII SK-
16	106-043-4-0-H3-CS	DNA	pBluescriptII SK-
17	110-007-1-0-C7-CS	DNA	pBluescriptII SK-
18	114-016-1-0-H8-CS	DNA	pBluescriptII SK-
19	116-004-3-0-A6-CS	DNA	pBluescriptII SK-
20	116-054-3-0-E6-CS	DNA	pBluescriptII SK-
21	116-055-1-0-A3-CS	DNA	pBluescriptII SK-
22	116-055-2-0-F7-CS	DNA	pBluescriptII SK-
23	116-088-4-0-A9-CS	DNA	pBluescriptII SK-
24	116-091-1-0-D9-CS	DNA	pBluescriptII SK-
25	116-110-2-0-F4-CS	DNA	pBluescriptII SK-
26	116-111-1-0-H9-CS	DNA	pBluescriptII SK-
27	116-111-4-0-B3-CS	DNA	pBluescriptII SK-
28	116-115-2-0-F8-CS	DNA	pBluescriptII SK-
29	116-119-3-0-H5-CS	DNA	pBluescriptII SK-
30	117-001-5-0-G3-CS	DNA	pBluescriptII SK-
31	145-25-3-0-B4-CS.cor	DNA	pBluescriptII SK-
32	145-25-3-0-B4-CS.fr	DNA	pBluescriptII SK-
33	145-56-3-0-D5-CS	DNA	pBluescriptII SK-
34	145-59-2-0-A7-CS	DNA	pBluescriptII SK-
35	157-15-4-0-B11-CS	DNA	pBluescriptII SK-
36	160-103-1-0-F11-CS	DNA	pBluescriptII SK-
37	160-37-2-0-H7-CS	DNA	pBluescriptII SK-
38	160-58-3-0-H3-CS	DNA	pBluescriptII SK-
39	160-75-4-0-A9-CS	DNA	pBluescriptII SK-
40	174-10-2-0-F8-CS	DNA	pPT
41	174-33-3-0-F6-CS	DNA	pPT
42	174-38-1-0-B6-CS	DNA	pPT
43	174-38-3-0-C9-CS	DNA	pPT
44	174-39-2-0-A3-CS	DNA	pPT
45	174-41-1-0-A6-CS	DNA	pPT
46	174-5-3-0-H7-CS	DNA	pPT
47	174-7-4-0-H1-CS	DNA	pPT
48	175-1-3-0-E5-CS.cor	DNA	pPT

49	175-1-3-0-E5-CS.fr	DNA	pPT
50	180-19-4-0-F4-CS	DNA	pBluescriptII SK-
51	181-10-1-0-D10-CS	DNA	pBluescriptII SK-
52	181-16-1-0-G7-CS	DNA	pBluescriptII SK-
53	181-16-2-0-A7-CS	DNA	pBluescriptII SK-
54	181-20-3-0-B5-CS	DNA	pBluescriptII SK-
55	181-3-3-0-B8-CS	DNA	pBluescriptII SK-
56	181-3-3-0-C9-CS	DNA	pBluescriptII SK-
57	182-1-2-0-D12-CS	DNA	pBluescriptII SK-
58	184-1-4-0-C11-CS	DNA	pBluescriptII SK-
59	184-4-1-0-A11-CS	DNA	pBluescriptII SK-
60	187-12-4-0-A8-CS	DNA	pBluescriptII SK-
61	187-2-2-0-A3-CS	DNA	pBluescriptII SK-
62	187-31-0-0-f12-CS	DNA	pBluescriptII SK-
63	187-34-0-0-l12-CS	DNA	pBluescriptII SK-
64	187-37-0-0-c10-CS	DNA	pBluescriptII SK-
65	187-38-0-0-l10-CS	DNA	pBluescriptII SK-
66	187-39-0-0-k12-CS	DNA	pBluescriptII SK-
67	187-41-0-0-i21-CS	DNA	pBluescriptII SK-
68	188-11-1-0-B3-CS	DNA	pBluescriptII SK-
69	188-18-4-0-A9-CS	DNA	pBluescriptII SK-
70	188-28-4-0-B12-CS.cor	DNA	pBluescriptII SK-
71	188-28-4-0-B12-CS.fr	DNA	pBluescriptII SK-
72	188-28-4-0-D4-CS	DNA	pBluescriptII SK-
73	188-41-1-0-B8-CS.cor	DNA	pBluescriptII SK-
74	188-41-1-0-B8-CS.fr	DNA	pBluescriptII SK-
75	188-45-1-0-D9-CS	DNA	pBluescriptII SK-
76	188-9-2-0-E1-CS	DNA	pBluescriptII SK-
77	105-079-3-0-A11-CS	DNA	pBluescriptII SK-
78	105-092-1-0-H7-CS	DNA	pBluescriptII SK-
79	105-141-4-0-H9-CS	DNA	pBluescriptII SK-
80	109-013-1-0-B9-CS	DNA	pBluescriptII SK-
81	110-008-4-0-D9-CS	DNA	pBluescriptII SK-
82	114-001-3-0-A2-CS	DNA	pBluescriptII SK-
83	114-028-2-0-C1-CS	DNA	pBluescriptII SK-
84	114-032-1-0-H10-CS	DNA	pBluescriptII SK-
85	114-043-2-0-A10-CS	DNA	pBluescriptII SK-
86	114-044-1-0-C5-CS	DNA	pBluescriptII SK-
87	116-003-3-0-D10-CS	DNA	pBluescriptII SK-
88	116-003-3-0-G12-CS	DNA	pBluescriptII SK-
89	116-011-2-0-F11-CS	DNA	pBluescriptII SK-
90	116-033-3-0-E4-CS	DNA	pBluescriptII SK-
91	116-041-4-0-B6-CS	DNA	pBluescriptII SK-
92	116-044-2-0-C4-CS	DNA	pBluescriptII SK-
93	116-075-1-0-E6-CS	DNA	pBluescriptII SK-
94	116-094-4-0-G5-CS	DNA	pBluescriptII SK-
95	117-005-3-0-F2-CS	DNA	pBluescriptII SK-
96	121-007-3-0-D9-CS	DNA	pBluescriptII SK-
97	145-91-3-0-D10-CS	DNA	pBluescriptII SK-
98	157-17-1-0-F4-CS	DNA	pBluescriptII SK-
99	160-11-3-0-G8-CS	DNA	pBluescriptII SK-
100	160-24-1-0-F12-CS	DNA	pBluescriptII SK-

101	160-24-2-0-E9-CS	DNA	pBluescriptII SK-
102	160-25-4-0-D2-CS	DNA	pBluescriptII SK-
103	160-31-3-0-A11-CS	DNA	pBluescriptII SK-
104	160-32-1-0-F6-CS	DNA	pBluescriptII SK-
105	160-37-1-0-A3-CS	DNA	pBluescriptII SK-
106	160-40-3-0-E9-CS	DNA	pBluescriptII SK-
107	160-58-3-0-E4-CS	DNA	pBluescriptII SK-
108	160-85-3-0-D4-CS	DNA	pBluescriptII SK-
109	160-95-3-0-A11-CS	DNA	pBluescriptII SK-
110	162-10-4-0-F9-CS.cor	DNA	pBluescriptII SK-
111	162-10-4-0-F9-CS.fr	DNA	pBluescriptII SK-
112	174-13-2-0-E4-CS	DNA	pPT
113	174-46-2-0-B11-CS	DNA	pPT
114	179-8-2-0-A6-CS	DNA	pBluescriptII SK-
115	180-22-3-0-B6-CS	DNA	pBluescriptII SK-
116	181-13-1-0-F7-CS	DNA	pBluescriptII SK-
117	181-15-4-0-F7-CS	DNA	pBluescriptII SK-
118	181-20-1-0-G7-CS	DNA	pBluescriptII SK-
119	184-15-3-0-D1-CS	DNA	pBluescriptII SK-
120	187-12-2-0-G11-CS	DNA	pBluescriptII SK-
121	187-2-2-0-A12-CS	DNA	pBluescriptII SK-
122	187-30-0-0-k23-CS	DNA	pBluescriptII SK-
123	187-36-0-0-e19-CS	DNA	pBluescriptII SK-
124	187-38-0-0-d22-CS	DNA	pBluescriptII SK-
125	187-39-0-0-b9-CS	DNA	pBluescriptII SK-
126	187-39-0-0-g6-CS	DNA	pBluescriptII SK-
127	187-45-0-0-l18-CS	DNA	pBluescriptII SK-
128	187-45-0-0-m21-CS	DNA	pBluescriptII SK-
129	187-45-0-0-n8-CS	DNA	pBluescriptII SK-
130	187-46-0-0-f23-CS	DNA	pBluescriptII SK-
131	187-5-1-0-A12-CS	DNA	pBluescriptII SK-
132	187-5-1-0-F6-CS	DNA	pBluescriptII SK-
133	187-5-2-0-B2-CS	DNA	pBluescriptII SK-
134	187-5-3-0-D5-CS	DNA	pBluescriptII SK-
135	187-51-0-0-f9-CS	DNA	pBluescriptII SK-
136	187-6-1-0-B9-CS	DNA	pBluescriptII SK-
137	187-6-4-0-C10-CS	DNA	pBluescriptII SK-
138	188-19-2-0-C8-CS	DNA	pBluescriptII SK-
139	188-22-4-0-G6-CS	DNA	pBluescriptII SK-
140	188-28-4-0-D11-CS	DNA	pBluescriptII SK-
141	188-29-1-0-E10-CS	DNA	pBluescriptII SK-
142	188-34-4-0-E5-CS	DNA	pBluescriptII SK-
143	188-9-3-0-A5-CS	DNA	pBluescriptII SK-
144	105-021-3-0-C3-CS	DNA	pBluescriptII SK-
145	105-037-4-0-H12-CS	DNA	pBluescriptII SK-
146	105-073-2-0-A7-CS	DNA	pBluescriptII SK-
147	109-002-4-0-C6-CS	DNA	pBluescriptII SK-
148	109-003-1-0-G4-CS	DNA	pBluescriptII SK-
149	116-118-4-0-A8-CS	DNA	pBluescriptII SK-
150	145-52-2-0-D12-CS	DNA	pBluescriptII SK-
151	145-7-2-0-G5-CS	DNA	pBluescriptII SK-
152	145-7-3-0-D3-CS	DNA	pBluescriptII SK-

153	157-17-2-0-C1-CS	DNA	pBluescriptII SK-
154	160-101-3-0-H2-CS	DNA	pBluescriptII SK-
155	160-12-1-0-D10-CS	DNA	pBluescriptII SK-
156	160-28-4-0-C4-CS	DNA	pBluescriptII SK-
157	160-31-3-0-E4-CS	DNA	pBluescriptII SK-
158	160-40-1-0-H4-CS	DNA	pBluescriptII SK-
159	160-54-1-0-F7-CS	DNA	pBluescriptII SK-
160	160-88-3-0-A8-CS.cor	DNA	pBluescriptII SK-
161	160-88-3-0-A8-CS.fr	DNA	pBluescriptII SK-
162	160-99-4-0-E4-CS	DNA	pBluescriptII SK-
163	161-5-4-0-B6-CS	DNA	pBluescriptII SK-
164	174-17-1-0-D6-CS	DNA	pPT
165	174-32-4-0-F8-CS	DNA	pPT
166	174-38-4-0-D11-CS	DNA	pPT
167	174-8-2-0-C10-CS	DNA	pPT
168	179-14-2-0-F11-CS	DNA	pBluescriptII SK-
169	179-9-4-0-B8-CS	DNA	pBluescriptII SK-
170	181-10-1-0-C9-CS	DNA	pBluescriptII SK-
171	187-5-3-0-C7-CS	DNA	pBluescriptII SK-
172	188-26-4-0-F5-CS	DNA	pBluescriptII SK-
173	188-27-3-0-G1-CS	DNA	pBluescriptII SK-
174	188-29-2-0-H1-CS	DNA	pBluescriptII SK-
175	188-31-1-0-E6-CS	DNA	pBluescriptII SK-
176	188-45-1-0-D3-CS	DNA	pBluescriptII SK-
177	188-5-1-0-H6-CS	DNA	pBluescriptII SK-
178	188-9-1-0-C10-CS	DNA	pBluescriptII SK-
179	105-016-3-0-C5-CS	DNA	pBluescriptII SK-
180	105-026-4-0-D9-CS	DNA	pBluescriptII SK-
181	105-053-2-0-D9-CS	DNA	pBluescriptII SK-
182	105-069-3-0-A11-CS	DNA	pBluescriptII SK-
183	105-076-4-0-F6-CS	DNA	pBluescriptII SK-
184	105-135-2-0-F9-CS	DNA	pBluescriptII SK-
185	106-023-4-0-F6-CS	DNA	pBluescriptII SK-
186	110-001-3-0-C11-CS	DNA	pBluescriptII SK-
187	110-002-3-0-F9-CS	DNA	pBluescriptII SK-
188	114-019-3-0-D9-CS	DNA	pBluescriptII SK-
189	114-029-1-0-C6-CS	DNA	pBluescriptII SK-
190	114-032-4-0-B1-CS	DNA	pBluescriptII SK-
191	114-070-2-0-H4-CS	DNA	pBluescriptII SK-
192	116-016-3-0-F11-CS	DNA	pBluescriptII SK-
193	116-022-4-0-G2-CS	DNA	pBluescriptII SK-
194	116-052-2-0-H8-CS	DNA	pBluescriptII SK-
195	116-053-4-0-B4-CS	DNA	pBluescriptII SK-
196	116-094-3-0-H2-CS	DNA	pBluescriptII SK-
197	116-112-4-0-C7-CS	DNA	pBluescriptII SK-
198	116-123-3-0-F12-CS	DNA	pBluescriptII SK-
199	123-008-1-0-C5-CS	DNA	pBluescriptII SK-
200	145-53-2-0-H8-CS	DNA	pBluescriptII SK-
201	145-57-2-0-C9-CS.cor	DNA	pBluescriptII SK-
202	145-57-2-0-C9-CS.fr	DNA	pBluescriptII SK-
203	145-7-3-0-B12-CS	DNA	pBluescriptII SK-
204	157-12-2-0-D1-CS	DNA	pBluescriptII SK-

205	157-16-2-0-D5-CS	DNA	pBluescriptII SK-
206	157-18-2-0-A7-CS	DNA	pBluescriptII SK-
207	160-103-1-0-B10-CS	DNA	pBluescriptII SK-
208	160-104-4-0-F3-CS	DNA	pBluescriptII SK-
209	160-22-2-0-D10-CS	DNA	pBluescriptII SK-
210	160-24-3-0-F12-CS	DNA	pBluescriptII SK-
211	160-3-2-0-H3-CS	DNA	pBluescriptII SK-
212	160-58-2-0-A2-CS	DNA	pBluescriptII SK-
213	160-73-1-0-B4-CS	DNA	pBluescriptII SK-
214	160-75-4-0-E6-CS	DNA	pBluescriptII SK-
215	160-97-3-0-E9-CS	DNA	pBluescriptII SK-
216	174-1-4-0-E9-CS	DNA	pPT
217	174-12-4-0-C2-CS	DNA	pPT
218	180-19-4-0-H2-CS	DNA	pBluescriptII SK-
219	181-10-4-0-G12-CS	DNA	pBluescriptII SK-
220	181-3-2-0-F6-CS	DNA	pBluescriptII SK-
221	181-4-4-0-A12-CS	DNA	pBluescriptII SK-
222	181-9-2-0-F12-CS.cor	DNA	pBluescriptII SK-
223	181-9-2-0-F12-CS.fr	DNA	pBluescriptII SK-
224	184-13-3-0-E11-CS	DNA	pBluescriptII SK-
225	184-4-2-0-D3-CS	DNA	pBluescriptII SK-
226	184-7-1-0-E7-CS	DNA	pBluescriptII SK-
227	184-8-4-0-G9-CS	DNA	pBluescriptII SK-
228	187-10-3-0-G9-CS	DNA	pBluescriptII SK-
229	187-32-0-0-m20-CS	DNA	pBluescriptII SK-
230	187-32-0-0-n21-CS.cor	DNA	pBluescriptII SK-
231	187-32-0-0-n21-CS.fr	DNA	pBluescriptII SK-
232	187-4-2-0-E6-CS	DNA	pBluescriptII SK-
233	187-40-0-0-i15-CS	DNA	pBluescriptII SK-
234	187-47-0-0-g24-CS	DNA	pBluescriptII SK-
235	187-9-3-0-A2-CS	DNA	pBluescriptII SK-
236	188-26-4-0-H1-CS	DNA	pBluescriptII SK-
237	188-35-3-0-G9-CS	DNA	pBluescriptII SK-
238	188-38-4-0-D8-CS	DNA	pBluescriptII SK-
239	188-41-1-0-E6-CS	DNA	pBluescriptII SK-
240	188-42-2-0-F3-CS.cor	DNA	pBluescriptII SK-
241	188-42-2-0-F3-CS.fr	DNA	pBluescriptII SK-
242	119-003-4-0-C2-CS	PRT	pBluescriptII SK-
243	105-016-1-0-D3-CS	PRT	pBluescriptII SK-
244	105-016-3-0-G10-CS	PRT	pBluescriptII SK-
245	105-026-1-0-A5-CS	PRT	pBluescriptII SK-
246	105-031-1-0-A11-CS	PRT	pBluescriptII SK-
247	105-031-2-0-D3-CS	PRT	pBluescriptII SK-
248	105-035-2-0-C6-CS	PRT	pBluescriptII SK-
249	105-037-2-0-H11-CS	PRT	pBluescriptII SK-
250	105-053-4-0-E8-CS	PRT	pBluescriptII SK-
251	105-074-3-0-H10-CS	PRT	pBluescriptII SK-
252	105-089-3-0-G10-CS	PRT	pBluescriptII SK-
253	105-095-2-0-G11-CS	PRT	pBluescriptII SK-
254	106-006-1-0-E3-CS	PRT	pBluescriptII SK-
255	106-037-1-0-E9-CS.cor	PRT	pBluescriptII SK-
256	106-037-1-0-E9-CS.fr	PRT	pBluescriptII SK-

257	106-043-4-0-H3-CS	PRT	pBluescriptII SK-
258	110-007-1-0-C7-CS	PRT	pBluescriptII SK-
259	114-016-1-0-H8-CS	PRT	pBluescriptII SK-
260	116-004-3-0-A6-CS	PRT	pBluescriptII SK-
261	116-054-3-0-E6-CS	PRT	pBluescriptII SK-
262	116-055-1-0-A3-CS	PRT	pBluescriptII SK-
263	116-055-2-0-F7-CS	PRT	pBluescriptII SK-
264	116-088-4-0-A9-CS	PRT	pBluescriptII SK-
265	116-091-1-0-D9-CS	PRT	pBluescriptII SK-
266	116-110-2-0-F4-CS	PRT	pBluescriptII SK-
267	116-111-1-0-H9-CS	PRT	pBluescriptII SK-
268	116-111-4-0-B3-CS	PRT	pBluescriptII SK-
269	116-115-2-0-F8-CS	PRT	pBluescriptII SK-
270	116-119-3-0-H5-CS	PRT	pBluescriptII SK-
271	117-001-5-0-G3-CS	PRT	pBluescriptII SK-
272	145-25-3-0-B4-CS.cor	PRT	pBluescriptII SK-
273	145-25-3-0-B4-CS.fr	PRT	pBluescriptII SK-
274	145-56-3-0-D5-CS	PRT	pBluescriptII SK-
275	145-59-2-0-A7-CS	PRT	pBluescriptII SK-
276	157-15-4-0-B11-CS	PRT	pBluescriptII SK-
277	160-103-1-0-F11-CS	PRT	pBluescriptII SK-
278	160-37-2-0-H7-CS	PRT	pBluescriptII SK-
279	160-58-3-0-H3-CS	PRT	pBluescriptII SK-
280	160-75-4-0-A9-CS	PRT	pBluescriptII SK-
281	174-10-2-0-F8-CS	PRT	pPT
282	174-33-3-0-F6-CS	PRT	pPT
283	174-38-1-0-B6-CS	PRT	pPT
284	174-38-3-0-C9-CS	PRT	pPT
285	174-39-2-0-A3-CS	PRT	pPT
286	174-41-1-0-A6-CS	PRT	pPT
287	174-5-3-0-H7-CS	PRT	pPT
288	174-7-4-0-H1-CS	PRT	pPT
289	175-1-3-0-E5-CS.cor	PRT	pPT
290	175-1-3-0-E5-CS.fr	PRT	pPT
291	180-19-4-0-F4-CS	PRT	pBluescriptII SK-
292	181-10-1-0-D10-CS	PRT	pBluescriptII SK-
293	181-16-1-0-G7-CS	PRT	pBluescriptII SK-
294	181-16-2-0-A7-CS	PRT	pBluescriptII SK-
295	181-20-3-0-B5-CS	PRT	pBluescriptII SK-
296	181-3-3-0-B8-CS	PRT	pBluescriptII SK-
297	181-3-3-0-C9-CS	PRT	pBluescriptII SK-
298	182-1-2-0-D12-CS	PRT	pBluescriptII SK-
299	184-1-4-0-C11-CS	PRT	pBluescriptII SK-
300	184-4-1-0-A11-CS	PRT	pBluescriptII SK-
301	187-12-4-0-A8-CS	PRT	pBluescriptII SK-
302	187-2-2-0-A3-CS	PRT	pBluescriptII SK-
303	187-31-0-0-f12-CS	PRT	pBluescriptII SK-
304	187-34-0-0-l12-CS	PRT	pBluescriptII SK-
305	187-37-0-0-c10-CS	PRT	pBluescriptII SK-
306	187-38-0-0-l10-CS	PRT	pBluescriptII SK-
307	187-39-0-0-k12-CS	PRT	pBluescriptII SK-
308	187-41-0-0-i21-CS	PRT	pBluescriptII SK-

309	188-11-1-0-B3-CS	PRT	pBluescriptII SK-
310	188-18-4-0-A9-CS	PRT	pBluescriptII SK-
311	188-28-4-0-B12-CS.cor	PRT	pBluescriptII SK-
312	188-28-4-0-B12-CS.fr	PRT	pBluescriptII SK-
313	188-28-4-0-D4-CS	PRT	pBluescriptII SK-
314	188-41-1-0-B8-CS.cor	PRT	pBluescriptII SK-
315	188-41-1-0-B8-CS.fr	PRT	pBluescriptII SK-
316	188-45-1-0-D9-CS	PRT	pBluescriptII SK-
317	188-9-2-0-E1-CS	PRT	pBluescriptII SK-
318	105-079-3-0-A11-CS	PRT	pBluescriptII SK-
319	105-092-1-0-H7-CS	PRT	pBluescriptII SK-
320	105-141-4-0-H9-CS	PRT	pBluescriptII SK-
321	109-013-1-0-B9-CS	PRT	pBluescriptII SK-
322	110-008-4-0-D9-CS	PRT	pBluescriptII SK-
323	114-001-3-0-A2-CS	PRT	pBluescriptII SK-
324	114-028-2-0-C1-CS	PRT	pBluescriptII SK-
325	114-032-1-0-H10-CS	PRT	pBluescriptII SK-
326	114-043-2-0-A10-CS	PRT	pBluescriptII SK-
327	114-044-1-0-C5-CS	PRT	pBluescriptII SK-
328	116-003-3-0-D10-CS	PRT	pBluescriptII SK-
329	116-003-3-0-G12-CS	PRT	pBluescriptII SK-
330	116-011-2-0-F11-CS	PRT	pBluescriptII SK-
331	116-033-3-0-E4-CS	PRT	pBluescriptII SK-
332	116-041-4-0-B6-CS	PRT	pBluescriptII SK-
333	116-044-2-0-C4-CS	PRT	pBluescriptII SK-
334	116-075-1-0-E6-CS	PRT	pBluescriptII SK-
335	116-094-4-0-G5-CS	PRT	pBluescriptII SK-
336	117-005-3-0-F2-CS	PRT	pBluescriptII SK-
337	121-007-3-0-D9-CS	PRT	pBluescriptII SK-
338	145-91-3-0-D10-CS	PRT	pBluescriptII SK-
339	157-17-1-0-F4-CS	PRT	pBluescriptII SK-
340	160-11-3-0-G8-CS	PRT	pBluescriptII SK-
341	160-24-1-0-F12-CS	PRT	pBluescriptII SK-
342	160-24-2-0-E9-CS	PRT	pBluescriptII SK-
343	160-25-4-0-D2-CS	PRT	pBluescriptII SK-
344	160-31-3-0-A11-CS	PRT	pBluescriptII SK-
345	160-32-1-0-F6-CS	PRT	pBluescriptII SK-
346	160-37-1-0-A3-CS	PRT	pBluescriptII SK-
347	160-40-3-0-E9-CS	PRT	pBluescriptII SK-
348	160-58-3-0-E4-CS	PRT	pBluescriptII SK-
349	160-85-3-0-D4-CS	PRT	pBluescriptII SK-
350	160-95-3-0-A11-CS	PRT	pBluescriptII SK-
351	162-10-4-0-F9-CS.cor	PRT	pBluescriptII SK-
352	162-10-4-0-F9-CS.fr	PRT	pBluescriptII SK-
353	174-13-2-0-E4-CS	PRT	pPT
354	174-46-2-0-B11-CS	PRT	pPT
355	179-8-2-0-A6-CS	PRT	pBluescriptII SK-
356	180-22-3-0-B6-CS	PRT	pBluescriptII SK-
357	181-13-1-0-F7-CS	PRT	pBluescriptII SK-
358	181-15-4-0-F7-CS	PRT	pBluescriptII SK-
359	181-20-1-0-G7-CS	PRT	pBluescriptII SK-
360	184-15-3-0-D1-CS	PRT	pBluescriptII SK-

361	187-12-2-0-G11-CS	PRT	pBluescriptII SK-
362	187-2-2-0-A12-CS	PRT	pBluescriptII SK-
363	187-30-0-0-k23-CS	PRT	pBluescriptII SK-
364	187-36-0-0-e19-CS	PRT	pBluescriptII SK-
365	187-38-0-0-d22-CS	PRT	pBluescriptII SK-
366	187-39-0-0-b9-CS	PRT	pBluescriptII SK-
367	187-39-0-0-g6-CS	PRT	pBluescriptII SK-
368	187-45-0-0-l18-CS	PRT	pBluescriptII SK-
369	187-45-0-0-m21-CS	PRT	pBluescriptII SK-
370	187-45-0-0-n8-CS	PRT	pBluescriptII SK-
371	187-46-0-0-f23-CS	PRT	pBluescriptII SK-
372	187-5-1-0-A12-CS	PRT	pBluescriptII SK-
373	187-5-1-0-F6-CS	PRT	pBluescriptII SK-
374	187-5-2-0-B2-CS	PRT	pBluescriptII SK-
375	187-5-3-0-D5-CS	PRT	pBluescriptII SK-
376	187-51-0-0-f9-CS	PRT	pBluescriptII SK-
377	187-6-1-0-B9-CS	PRT	pBluescriptII SK-
378	187-6-4-0-C10-CS	PRT	pBluescriptII SK-
379	188-19-2-0-C8-CS	PRT	pBluescriptII SK-
380	188-22-4-0-G6-CS	PRT	pBluescriptII SK-
381	188-28-4-0-D11-CS	PRT	pBluescriptII SK-
382	188-29-1-0-E10-CS	PRT	pBluescriptII SK-
383	188-34-4-0-E5-CS	PRT	pBluescriptII SK-
384	188-9-3-0-A5-CS	PRT	pBluescriptII SK-
385	105-021-3-0-C3-CS	PRT	pBluescriptII SK-
386	105-037-4-0-H12-CS	PRT	pBluescriptII SK-
387	105-073-2-0-A7-CS	PRT	pBluescriptII SK-
388	109-002-4-0-C6-CS	PRT	pBluescriptII SK-
389	109-003-1-0-G4-CS	PRT	pBluescriptII SK-
390	116-118-4-0-A8-CS	PRT	pBluescriptII SK-
391	145-52-2-0-D12-CS	PRT	pBluescriptII SK-
392	145-7-2-0-G5-CS	PRT	pBluescriptII SK-
393	145-7-3-0-D3-CS	PRT	pBluescriptII SK-
394	157-17-2-0-C1-CS	PRT	pBluescriptII SK-
395	160-101-3-0-H2-CS	PRT	pBluescriptII SK-
396	160-12-1-0-D10-CS	PRT	pBluescriptII SK-
397	160-28-4-0-C4-CS	PRT	pBluescriptII SK-
398	160-31-3-0-E4-CS	PRT	pBluescriptII SK-
399	160-40-1-0-H4-CS	PRT	pBluescriptII SK-
400	160-54-1-0-F7-CS	PRT	pBluescriptII SK-
401	160-88-3-0-A8-CS.cor	PRT	pBluescriptII SK-
402	160-88-3-0-A8-CS.fr	PRT	pBluescriptII SK-
403	160-99-4-0-E4-CS	PRT	pBluescriptII SK-
404	161-5-4-0-B6-CS	PRT	pBluescriptII SK-
405	174-17-1-0-D6-CS	PRT	pPT
406	174-32-4-0-F8-CS	PRT	pPT
407	174-38-4-0-D11-CS	PRT	pPT
408	174-8-2-0-C10-CS	PRT	pPT
409	179-14-2-0-F11-CS	PRT	pBluescriptII SK-
410	179-9-4-0-B8-CS	PRT	pBluescriptII SK-
411	181-10-1-0-C9-CS	PRT	pBluescriptII SK-
412	187-5-3-0-C7-CS	PRT	pBluescriptII SK-

413	188-26-4-0-F5-CS	PRT	pBluescriptII SK-
414	188-27-3-0-G1-CS	PRT	pBluescriptII SK-
415	188-29-2-0-H1-CS	PRT	pBluescriptII SK-
416	188-31-1-0-E6-CS	PRT	pBluescriptII SK-
417	188-45-1-0-D3-CS	PRT	pBluescriptII SK-
418	188-5-1-0-H6-CS	PRT	pBluescriptII SK-
419	188-9-1-0-C10-CS	PRT	pBluescriptII SK-
420	105-016-3-0-C5-CS	PRT	pBluescriptII SK-
421	105-026-4-0-D9-CS	PRT	pBluescriptII SK-
422	105-053-2-0-D9-CS	PRT	pBluescriptII SK-
423	105-069-3-0-A11-CS	PRT	pBluescriptII SK-
424	105-076-4-0-F6-CS	PRT	pBluescriptII SK-
425	105-135-2-0-F9-CS	PRT	pBluescriptII SK-
426	106-023-4-0-F6-CS	PRT	pBluescriptII SK-
427	110-001-3-0-C11-CS	PRT	pBluescriptII SK-
428	110-002-3-0-F9-CS	PRT	pBluescriptII SK-
429	114-019-3-0-D9-CS	PRT	pBluescriptII SK-
430	114-029-1-0-C6-CS	PRT	pBluescriptII SK-
431	114-032-4-0-B1-CS	PRT	pBluescriptII SK-
432	114-070-2-0-H4-CS	PRT	pBluescriptII SK-
433	116-016-3-0-F11-CS	PRT	pBluescriptII SK-
434	116-022-4-0-G2-CS	PRT	pBluescriptII SK-
435	116-052-2-0-H8-CS	PRT	pBluescriptII SK-
436	116-053-4-0-B4-CS	PRT	pBluescriptII SK-
437	116-094-3-0-H2-CS	PRT	pBluescriptII SK-
438	116-112-4-0-C7-CS	PRT	pBluescriptII SK-
439	116-123-3-0-F12-CS	PRT	pBluescriptII SK-
440	123-008-1-0-C5-CS	PRT	pBluescriptII SK-
441	145-53-2-0-H8-CS	PRT	pBluescriptII SK-
442	145-57-2-0-C9-CS.cor	PRT	pBluescriptII SK-
443	145-57-2-0-C9-CS.fr	PRT	pBluescriptII SK-
444	145-7-3-0-B12-CS	PRT	pBluescriptII SK-
445	157-12-2-0-D1-CS	PRT	pBluescriptII SK-
446	157-16-2-0-D5-CS	PRT	pBluescriptII SK-
447	157-18-2-0-A7-CS	PRT	pBluescriptII SK-
448	160-103-1-0-B10-CS	PRT	pBluescriptII SK-
449	160-104-4-0-F3-CS	PRT	pBluescriptII SK-
450	160-22-2-0-D10-CS	PRT	pBluescriptII SK-
451	160-24-3-0-F12-CS	PRT	pBluescriptII SK-
452	160-3-2-0-H3-CS	PRT	pBluescriptII SK-
453	160-58-2-0-A2-CS	PRT	pBluescriptII SK-
454	160-73-1-0-B4-CS	PRT	pBluescriptII SK-
455	160-75-4-0-E6-CS	PRT	pBluescriptII SK-
456	160-97-3-0-E9-CS	PRT	pBluescriptII SK-
457	174-1-4-0-E9-CS	PRT	pPT
458	174-12-4-0-C2-CS	PRT	pPT
459	180-19-4-0-H2-CS	PRT	pBluescriptII SK-
460	181-10-4-0-G12-CS	PRT	pBluescriptII SK-
461	181-3-2-0-F6-CS	PRT	pBluescriptII SK-
462	181-4-4-0-A12-CS	PRT	pBluescriptII SK-
463	181-9-2-0-F12-CS.cor	PRT	pBluescriptII SK-
464	181-9-2-0-F12-CS.fr	PRT	pBluescriptII SK-

465	184-13-3-0-E11-CS	PRT	pBluescriptII SK-
466	184-4-2-0-D3-CS	PRT	pBluescriptII SK-
467	184-7-1-0-E7-CS	PRT	pBluescriptII SK-
468	184-8-4-0-G9-CS	PRT	pBluescriptII SK-
469	187-10-3-0-G9-CS	PRT	pBluescriptII SK-
470	187-32-0-0-m20-CS	PRT	pBluescriptII SK-
471	187-32-0-0-n21-CS.cor	PRT	pBluescriptII SK-
472	187-32-0-0-n21-CS.fr	PRT	pBluescriptII SK-
473	187-4-2-0-E6-CS	PRT	pBluescriptII SK-
474	187-40-0-0-i15-CS	PRT	pBluescriptII SK-
475	187-47-0-0-g24-CS	PRT	pBluescriptII SK-
476	187-9-3-0-A2-CS	PRT	pBluescriptII SK-
477	188-26-4-0-H1-CS	PRT	pBluescriptII SK-
478	188-35-3-0-G9-CS	PRT	pBluescriptII SK-
479	188-38-4-0-D8-CS	PRT	pBluescriptII SK-
480	188-41-1-0-E6-CS	PRT	pBluescriptII SK-
481	188-42-2-0-F3-CS.cor	PRT	pBluescriptII SK-
482	188-42-2-0-F3-CS.fr	PRT	pBluescriptII SK-

Table II

Seq Id No	Full coding sequence	Signal sequence	Coding sequence for mature protein	Polyadenylation signal	Polyadenylation site
1	[169-1692]	[169-249]	[250-1692]	[2126-2131]	[2152-2201]
2	[148-1140]	[148-240]	[241-1140]	[1592-1597]	[1615-1631]
3	[85-906]	[85-135]	[136-906]	[1159-1164]	[1184-1245]
4	[31-1248]	[31-135]	[136-1248]	None detected	[1607-1623]
5	[72-143]	[72-119]	[120-143]	[1416-1421]	[1438-1454]
6	[111-1154]	[111-197]	[198-1154]	[1602-1607]	[1623-1639]
7	[66-1256]	[66-173]	[174-1256]	None detected	[1752-1768]
8	[190-1398]	[190-252]	[253-1398]	[1470-1475]	[1494-1510]
9	[78-410]	[78-155]	[156-410]	None detected	[866-882]
10	[84-299]	[84-134]	[135-299]	[1814-1819]	[1833-1849]
11	[55-468]	[55-99]	[100-468]	[531-536]	[549-565]
12	[152-475]	[152-244]	[245-475]	[1623-1628]	[1647-1663]
13	[112-552]	[112-183]	[184-552]	[706-711]	[729-744]
14	[101-1243]	[101-199]	[200-1243]	[1720-1725]	[1745-1759]
15	[101-517]	[101-199]	[200-517]	[1716-1721]	[1741-1755]
16	[59-853]	[59-100]	[101-853]	[894-899]	[922-936]
17	[73-672]	[73-132]	[133-672]	[689-694]	[711-747]
18	[94-1275]	[94-210]	[211-1275]	[1849-1854]	[1870-1884]
19	[42-515]	[42-92]	[93-515]	[649-654]	[677-691]
20	[271-969]	[271-366]	[367-969]	[1093-1098]	[1124-1138]
21	[76-276]	[76-135]	[136-276]	[436-441]	[455-468]
22	[6-287]	[6-80]	[81-287]	[684-689]	[706-720]
23	[171-692]	[171-227]	[228-692]	[691-696]	[713-727]
24	[137-454]	[137-187]	[188-454]	[440-445]	[456-470]
25	[238-609]	[238-291]	[292-609]	[948-953]	[973-987]
26	[80-862]	[80-127]	[128-862]	[875-880]	[894-908]
27	[83-310]	[83-157]	[158-310]	[725-730]	[748-762]
28	[310-906]	[310-357]	[358-906]	[1071-1076]	[1088-1102]
29	[24-287]	[24-131]	[132-287]	[405-410]	[422-436]
30	[132-1574]	[132-206]	[207-1574]	[1899-1904]	[1923-1938]
31	[117-545]	[117-245]	[246-545]	None detected	[1100-1116]
32	[117-362]	none detected	[117-362]	None detected	[1098-1114]
33	[144-1262]	[144-224]	[225-1262]	[2035-2040]	[2056-2072]
34	[35-316]	[35-109]	[110-316]	None detected	[393-409]
35	[177-767]	[177-236]	[237-767]	None detected	[822-836]
36	[208-1239]	[208-294]	[295-1239]	None detected	[1307-1323]
37	[60-1682]	[60-143]	[144-1682]	None detected	[1929-1945]
38	[198-998]	[198-269]	[270-998]	[1292-1297]	[1315-1330]
39	[505-1590]	[505-624]	[625-1590]	[2089-2094]	[2108-2124]
40	[84-326]	[84-146]	[147-326]	[1122-1127]	[1142-1159]
41	[56-1678]	[56-139]	[140-1678]	None detected	[1936-1953]

42	[119-1522]	[119-181]	[182-1522]	None detected	[1671-1688]
43	[334-1551]	[334-426]	[427-1551]	None detected	[1925-1942]
44	[72-986]	[72-149]	[150-986]	[1608-1613]	[1640-1657]
45	[157-1482]	[157-219]	[220-1482]	None detected	[1716-1733]
46	[195-1052]	[195-338]	[339-1052]	None detected	[1854-1871]
47	[217-1410]	[217-279]	[280-1410]	[1482-1487]	[1507-1536]
48	[103-492]	[103-162]	[163-492]	[794-799]	[815-832]
49	[234-491]	[234-293]	[294-491]	[793-798]	[814-831]
50	[180-800]	[180-248]	[249-800]	[880-885]	[901-917]
51	[140-472]	[140-211]	[212-472]	None detected	[605-621]
52	[68-484]	[68-112]	[113-484]	None detected	[657-673]
53	[38-517]	[38-118]	[119-517]	[861-866]	[885-897]
54	[92-634]	[92-139]	[140-634]	None detected	None detected
55	[27-767]	[27-80]	[81-767]	None detected	[1031-1047]
56	[4-399]	[4-126]	[127-399]	[891-896]	[909-923]
57	[127-879]	[127-198]	[199-879]	None detected	[1224-1240]
58	[156-566]	[156-221]	[222-566]	[870-875]	[888-902]
59	[35-1657]	[35-118]	[119-1657]	None detected	[1955-1969]
60	[77-937]	[77-127]	[128-937]	[1098-1103]	[1116-1132]
61	[9-503]	[9-113]	[114-503]	[594-599]	[615-631]
62	[21-464]	[21-95]	[96-464]	[650-655]	[692-722]
63	[178-1050]	[178-279]	[280-1050]	[1400-1405]	[1426-1442]
64	[32-274]	[32-178]	[179-274]	[756-761]	[779-795]
65	[222-920]	[222-311]	[312-920]	[1191-1196]	[1220-1236]
66	[101-355]	[101-160]	[161-355]	[772-777]	[788-881]
67	[173-487]	[173-301]	[302-487]	[486-491]	[508-524]
68	[210-1082]	[210-311]	[312-1082]	[1432-1437]	[1456-1472]
69	[172-1449]	[172-255]	[256-1449]	None detected	[1721-1737]
70	[30-1427]	[30-77]	[78-1427]	[1594-1599]	[1621-1637]
71	[30-1175]	[30-77]	[78-1175]	[1593-1598]	[1620-1636]
72	[66-839]	[66-173]	[174-839]	None detected	[1742-1758]
73	[64-903]	[64-162]	[163-903]	[1612-1617]	[1631-1647]
74	[64-585]	[64-162]	[163-585]	[1611-1616]	[1630-1646]
75	[274-753]	[274-324]	[325-753]	[1931-1936]	[1947-1963]
76	[191-1468]	[191-274]	[275-1468]	None detected	[1741-1757]
77	[48-950]	[48-107]	[108-950]	[1983-1988]	[2011-2027]
78	[156-512]	[156-206]	[207-512]	[1831-1836]	[1864-1880]
79	[67-351]	[67-183]	[184-351]	None detected	[568-584]
80	[259-831]	[259-375]	[376-831]	None detected	[1337-1351]
81	[111-377]	[111-233]	[234-377]	[689-694]	[706-720]
82	[223-432]	[223-336]	[337-432]	[986-991]	[1015-1029]
83	[769-1272]	[769-843]	[844-1272]	None detected	[1774-1788]
84	[30-527]	[30-74]	[75-527]	[738-743]	[756-805]
85	[39-506]	[39-83]	[84-506]	None detected	[800-814]
86	[115-429]	[115-210]	[211-429]	[565-570]	[584-598]
87	[332-574]	[332-412]	[413-574]	None detected	[630-699]

88	[133-417]	[133-213]	[214-417]	[876-881]	[891-905]
89	[113-364]	[113-172]	[173-364]	None detected	[500-514]
90	[9-380]	[9-104]	[105-380]	[483-488]	[504-518]
91	[155-340]	[155-292]	[293-340]	[728-733]	[754-808]
92	[185-634]	[185-253]	[254-634]	[704-709]	[723-737]
93	[53-646]	[53-91]	[92-646]	[694-699]	[714-728]
94	[247-510]	[247-318]	[319-510]	[544-549]	[568-582]
95	[143-592]	[143-277]	[278-592]	[1877-1882]	[1898-1913]
96	[33-458]	[33-89]	[90-458]	[637-642]	[654-670]
97	[1-336]	[1-81]	[82-336]	[900-905]	[923-939]
98	[174-443]	[174-269]	[270-443]	[629-634]	[647-661]
99	[282-521]	[282-386]	[387-521]	[600-605]	[631-647]
100	[251-643]	[251-295]	[296-643]	None detected	[990-1006]
101	[179-475]	[179-295]	[296-475]	[995-1000]	[1015-1059]
102	[34-327]	[34-162]	[163-327]	[466-471]	[498-514]
103	[303-953]	[303-359]	[360-953]	[1124-1129]	[1142-1158]
104	[97-645]	[97-156]	[157-645]	[1524-1529]	[1547-1563]
105	[80-820]	[80-118]	[119-820]	[1587-1592]	[1606-1621]
106	[77-388]	[77-217]	[218-388]	[524-529]	[541-557]
107	[139-513]	[139-201]	[202-513]	[566-571]	[584-600]
108	[81-986]	[81-134]	[135-986]	[1092-1097]	[1113-1129]
109	[266-586]	[266-307]	[308-586]	[745-750]	[762-778]
110	[59-745]	[59-160]	[161-745]	None detected	[1285-1301]
111	[59-676]	[59-160]	[161-676]	None detected	[1284-1300]
112	[15-278]	[15-146]	[147-278]	[1580-1585]	[1600-1617]
113	[167-619]	[167-262]	[263-619]	[1598-1603]	[1617-1634]
114	[223-417]	[223-270]	[271-417]	[655-660]	[677-693]
115	[166-732]	[166-237]	[238-732]	[753-758]	[768-784]
116	[75-623]	[75-215]	[216-623]	[767-772]	[788-804]
117	[30-335]	[30-71]	[72-335]	[450-455]	[468-484]
118	[21-752]	[21-107]	[108-752]	None detected	[970-985]
119	[185-715]	[185-253]	[254-715]	[785-790]	[814-839]
120	[54-527]	[54-116]	[117-527]	[545-550]	[567-583]
121	[129-686]	[129-185]	[186-686]	[989-994]	[1008-1024]
122	[165-614]	[165-305]	[306-614]	[719-724]	[744-760]
123	[192-476]	[192-326]	[327-476]	[555-560]	[578-594]
124	[16-297]	[16-93]	[94-297]	None detected	[543-559]
125	[216-635]	[216-335]	[336-635]	[717-722]	[728-744]
126	[164-280]	[164-268]	[269-280]	[789-794]	[809-824]
127	[68-301]	[68-190]	[191-301]	[485-490]	[510-526]
128	[179-427]	[179-298]	[299-427]	[579-584]	[602-618]
129	[22-297]	[22-66]	[67-297]	[742-747]	[760-776]
130	[9-845]	[9-134]	[135-845]	[964-969]	[983-998]
131	[27-578]	[27-119]	[120-578]	[742-747]	[763-779]
132	[408-710]	[408-533]	[534-710]	[985-990]	[1009-1025]
133	[247-501]	[247-306]	[307-501]	None detected	[592-607]

134	[333-602]	[333-416]	[417-602]	None detected	[761-774]
135	[110-376]	[110-208]	[209-376]	[582-587]	[601-611]
136	[22-417]	[22-66]	[67-417]	[888-893]	[909-925]
137	[62-367]	[62-103]	[104-367]	[638-643]	[658-674]
138	[107-1618]	[107-178]	[179-1618]	[1688-1693]	[1709-1725]
139	[16-471]	[16-93]	[94-471]	None detected	[1458-1474]
140	[222-374]	[222-299]	[300-374]	None detected	[637-653]
141	[59-274]	[59-127]	[128-274]	[1452-1457]	[1474-1490]
142	[158-442]	[158-301]	[302-442]	[621-626]	[645-661]
143	[5-454]	[5-64]	[65-454]	[1745-1750]	[1773-1789]
144	[241-1302]	none detected	[241-1302]	[1968-1973]	[1990-2006]
145	[15-635]	none detected	[15-635]	[1057-1062]	[1080-1096]
146	[109-738]	none detected	[109-738]	[1633-1638]	[1650-1666]
147	[21-1145]	none detected	[21-1145]	[1648-1653]	[1666-1687]
148	[70-1596]	none detected	[70-1596]	[1712-1717]	[1733-1747]
149	[129-362]	none detected	[129-362]	[597-602]	[626-658]
150	[109-594]	none detected	[109-594]	[1999-2004]	[2029-2045]
151	[150-587]	none detected	[150-587]	None detected	[772-788]
152	[173-847]	none detected	[173-847]	[1894-1899]	[1915-1931]
153	[100-441]	none detected	[100-441]	[479-484]	[500-514]
154	[32-1132]	none detected	[32-1132]	None detected	[1167-1183]
155	[160-996]	none detected	[160-996]	[1504-1509]	[1529-1545]
156	[11-529]	none detected	[11-529]	[1042-1047]	[1053-1068]
157	[135-749]	none detected	[135-749]	[1055-1060]	[1081-1097]
158	[98-637]	none detected	[98-637]	[862-867]	[878-894]
159	[221-670]	none detected	[221-670]	[669-674]	[688-703]
160	[165-674]	none detected	[165-674]	[808-813]	[833-849]
161	[165-671]	none detected	[165-671]	[805-810]	[830-846]
162	[28-1128]	none detected	[28-1128]	[1121-1126]	[1159-1176]
163	[135-194]	none detected	[135-194]	[1050-1055]	[1068-1084]
164	[173-847]	none detected	[173-847]	[1757-1762]	[1776-1793]
165	[8-1141]	none detected	[8-1141]	None detected	[1832-1849]
166	[136-264]	none detected	[136-264]	[1720-1725]	[1731-1748]
167	[14-1048]	none detected	[14-1048]	[1234-1239]	[1258-1275]
168	[70-777]	none detected	[70-777]	[987-992]	[1007-1023]
169	[38-400]	none detected	[38-400]	[1043-1048]	[1069-1085]
170	[63-572]	none detected	[63-572]	[750-755]	[767-776]
171	[160-867]	none detected	[160-867]	[1178-1183]	[1203-1219]
172	[68-640]	none detected	[68-640]	None detected	[1471-1487]
173	[132-1298]	none detected	[132-1298]	[1873-1878]	[1899-1915]
174	[259-1701]	none detected	[259-1701]	None detected	[1974-1990]
175	[213-1274]	none detected	[213-1274]	[1940-1945]	[1955-1971]
176	[68-127]	none detected	[68-127]	None detected	[1597-1613]
177	[65-1024]	none detected	[65-1024]	[1291-1296]	[1315-1361]
178	[109-585]	none detected	[109-585]	[1059-1064]	[1082-1113]
179	[29-577]	none detected	[29-577]	[1917-1922]	[1944-1960]

180	[23-451]	none detected	[23-451]	[1405-1410]	[1427-1443]
181	[232-450]	none detected	[232-450]	None detected	[589-605]
182	[758-1183]	none detected	[758-1183]	None detected	[1708-1724]
183	[486-932]	none detected	[486-932]	None detected	[1670-1686]
184	[80-304]	none detected	[80-304]	None detected	[452-463]
185	[188-691]	none detected	[188-691]	[707-712]	[727-773]
186	[94-573]	none detected	[94-573]	None detected	[739-753]
187	[181-462]	none detected	[181-462]	None detected	[740-754]
188	[6-290]	none detected	[6-290]	None detected	[971-998]
189	[115-411]	none detected	[115-411]	[573-578]	[591-605]
190	[3-368]	none detected	[3-368]	[481-486]	[511-526]
191	[174-527]	none detected	[174-527]	[878-883]	[896-910]
192	[57-203]	none detected	[57-203]	[579-584]	[599-668]
193	[68-334]	none detected	[68-334]	[562-567]	[583-637]
194	[183-443]	none detected	[183-443]	[670-675]	[692-706]
195	[94-228]	none detected	[94-228]	None detected	[656-670]
196	[133-327]	none detected	[133-327]	[465-470]	[496-510]
197	[22-357]	none detected	[22-357]	None detected	[486-500]
198	[4-333]	none detected	[4-333]	[633-638]	[653-667]
199	[1-363]	none detected	[1-363]	[474-479]	[498-514]
200	[41-337]	none detected	[41-337]	None detected	[401-462]
201	[1-551]	none detected	[1-551]	None detected	[535-551]
202	[34-315]	none detected	[34-315]	None detected	[534-550]
203	[1-315]	none detected	[1-315]	[371-376]	[392-408]
204	[94-582]	none detected	[94-582]	None detected	[651-665]
205	[540-923]	none detected	[540-923]	None detected	[994-1008]
206	[77-364]	none detected	[77-364]	[367-372]	[391-455]
207	[65-544]	none detected	[65-544]	[710-715]	[733-749]
208	[117-467]	none detected	[117-467]	[557-562]	[578-594]
209	[893-1897]	none detected	[893-1897]	[2066-2071]	[2082-2098]
210	[85-342]	none detected	[85-342]	None detected	[412-428]
211	[155-433]	none detected	[155-433]	[713-718]	[735-769]
212	[63-386]	none detected	[63-386]	[878-883]	[898-914]
213	[460-1290]	none detected	[460-1290]	[1449-1454]	[1473-1489]
214	[21-539]	none detected	[21-539]	[741-746]	[760-776]
215	[34-1143]	none detected	[34-1143]	[1375-1380]	[1397-1412]
216	[6-1184]	none detected	[6-1184]	[1735-1740]	[1744-1773]
217	[29-376]	none detected	[29-376]	None detected	[1184-1251]
218	[78-566]	none detected	[78-566]	[858-863]	[878-894]
219	[16-705]	none detected	[16-705]	[868-873]	[894-910]
220	[103-405]	none detected	[103-405]	[482-487]	[503-519]
221	[72-350]	none detected	[72-350]	[593-598]	[616-632]
222	[38-436]	none detected	[38-436]	None detected	[636-652]
223	[38-322]	none detected	[38-322]	None detected	[634-650]
224	[202-480]	none detected	[202-480]	[472-477]	[488-502]
225	[171-1670]	none detected	[171-1670]	[1706-1711]	[1725-1739]

226	[199-618]	none detected	[199-618]	[626-631]	[643-657]
227	[182-481]	none detected	[182-481]	None detected	[874-888]
228	[161-517]	none detected	[161-517]	None detected	[701-716]
229	[86-505]	none detected	[86-505]	[618-623]	[638-654]
230	[56-382]	none detected	[56-382]	[598-603]	[619-635]
231	[56-355]	none detected	[56-355]	[597-602]	[618-634]
232	[76-498]	none detected	[76-498]	[546-551]	[567-583]
233	[199-600]	none detected	[199-600]	[705-710]	[737-753]
234	[211-612]	none detected	[211-612]	[717-722]	[746-762]
235	[5-259]	none detected	[5-259]	[502-507]	[521-537]
236	[23-370]	none detected	[23-370]	[956-961]	[978-994]
237	[41-352]	none detected	[41-352]	None detected	[646-662]
238	[3-1319]	none detected	[3-1319]	[1791-1796]	[1813-1829]
239	[421-768]	none detected	[421-768]	[1045-1050]	[1067-1083]
240	[78-590]	none detected	[78-590]	None detected	[1815-1831]
241	[78-608]	none detected	[78-608]	None detected	[1814-1830]

Table III

List of variants
92;119
14;15
110;111
69;174;76
2;12
172;176;177
150;152;164;166
154;162
77;143
34;62
230;231
63;68
8;47
48;49;66
7;72
160;161
144;175
17;21
31;32
5;6
3;10
96;121
37;41;59
70;71
19;24
186;195;204
73;74
240;241
221;235
222;223
42;45
157;163
190;229
117;137
122;233;234
201;202
80;139

Table IV

Seq Id No	Preferentially excluded fragments
1	192..235;2099..2201
2	174..225;1605..1631
3	1111..1245
4	1590..1598;1607..1623
5	1385..1453
6	1571..1639
7	1732..1768
8	1494..1510
9	570..882
10	1176..1218;1710..1742;1833..1849
11	219..253;455..565
12	178..229;1636..1663
13	729..744
14	790..827;1735..1759
15	788..825;1731..1755
16	922..936
17	668..747
18	1870..1884
19	677..691
20	1124..1138
21	450..468
22	393..411;706..720
23	713..727
24	456..470
25	876..928;973..987
26	894..908
27	748..762
28	1088..1102
29	422..436
30	1879..1918;1923..1938
31	774..1116
32	772..1114
33	2056..2072
34	393..409
35	784..836
36	544..551;1307..1323
37	1867..1874;1929..1945
38	1315..1330
39	2108..2124
40	413..421;1116..1159
41	1863..1870;1936..1953
42	1623..1688

43	1895..1942
44	1640..1657
45	1661..1733
46	1555..1871
47	1507..1523
48	541..832
49	540..831
50	901..917
51	2..10;605..621
52	585..673
53	885..897
54	4..13;761..1101
55	1031..1047
56	873..905;907..923
57	1224..1240
58	861..902
59	1842..1849;1955..1969
60	1116..1132
61	15..46;615..631
62	651..722
63	1426..1442
64	739..795
65	1220..1236
66	520..881
67	413..524
68	1444..1472
69	1721..1737
70	1621..1637
71	1620..1636
72	777..784;1742..1758
73	1631..1647
74	1630..1646
75	1947..1963
76	1741..1757
77	1561..1913;2011..2027
78	727..819;880..894;901..1280;1841..1880
79	418..584
80	331..353;844..1214;1337..1351
81	706..720
82	639..713;1008..1029
83	1454..1788
84	712..805
85	800..814
86	584..598
87	122..308;593..699
88	855..905

89	500..514
90	81..101;198..205;504..518
91	650..808
92	128..201;723..737
93	714..728
94	568..582
95	1761..1773;1898..1913
96	654..670
97	883..938
98	616..661
99	631..647
100	853..1006
101	537..544;949..1059
102	498..514
103	1142..1158
104	1524..1563
105	1230..1259;1606..1621
106	505..557
107	584..600
108	378..385;1113..1129
109	729..778
110	992..1301
111	991..1300
112	1131..1139;1569..1617
113	1526..1634
114	457..509;677..693
115	768..784
116	360..670;788..804
117	435..484
118	433..452;764..985
119	128..201;801..839
120	554..564;567..583
121	872..908;1008..1024
122	744..760
123	578..594
124	94..102;248..559
125	728..744
126	809..824
127	510..526
128	602..618
129	472..553;569..776
130	983..998
131	396..468;763..779
132	478..532;1009..1025
133	592..607
134	761..774

135	556..563;601..611
136	887..919
137	658..674
138	1651..1725
139	49..71;988..1358;1458..1474
140	324..653
141	720..730;1449..1490
142	44..119;498..505;578..585;645..661
143	1322..1666;1773..1789
144	1828..1897;1919..1968;1990..2006
145	936..955;1060..1096
146	778..827;1650..1666
147	1170..1207;1647..1687
148	1733..1747
149	579..658
150	1432..1440;1728..1778;2004..2045
151	772..788
152	1496..1504;1792..1842;1915..1931
153	500..514
154	1167..1183
155	1529..1545
156	703..1068
157	873..881;1081..1097
158	878..894
159	688..703
160	833..849
161	830..846
162	1159..1176
163	869..876;1068..1084
164	1444..1463;1496..1504;1743..1793
165	1233..1319;1697..1849
166	1407..1426;1459..1467;1694..1748
167	1258..1275
168	84..129;1002..1023
169	436..472;596..604;673..689;732..954;995..1085
170	767..776
171	1203..1219
172	1411..1487
173	1861..1915
174	1974..1990
175	1800..1869;1891..1940;1955..1971
176	1597..1613
177	186..212;1277..1361
178	930..978;1002..1113
179	951..1000;1364..1533;1944..1960
180	1427..1443

181	107..181;276..311;449..605
182	1143..1450;1677..1724
183	1..251;648..655;1347..1686
184	447..463
185	150..159;623..773
186	340..476;739..753
187	740..754
188	307..315;668..998
189	118..125;529..536;591..605
190	492..526
191	872..910
192	525..668
193	91..135;461..637
194	392..458;551..671;692..706
195	656..670
196	283..379;458..466;496..510
197	1..96;483..500
198	625..667
199	474..513
200	370..462
201	535..551
202	534..550
203	374..408
204	651..665
205	994..1008
206	348..455
207	733..749
208	1..49;578..594
209	2082..2098
210	412..428
211	689..769
212	898..914
213	1266..1489
214	760..776
215	1304..1311;1383..1412
216	648..691;1711..1773
217	644..856;910..1251
218	878..894
219	894..910
220	503..519
221	616..632
222	636..652
223	634..650
224	50..57;488..502
225	534..577;1725..1739
226	643..657

227	1..84;874..888
228	701..716
229	638..654
230	263..573;619..635
231	263..573;619..635
232	567..583
233	737..753
234	746..762
235	499..537
236	905..912;944..994
237	348..662
239	829..1083
240	1508..1831
241	1507..1830

Table Va

Seq Id No	Preferentially excluded fragments	Preferentially included fragments
1	[1-540];[556-615];[2061-2096];[2098-2201]	[541-555];[616-2060];[2097-2097]
2	[1-511];[533-619];[621-690];[730-1132]	[512-532];[620-620];[691-729];[1133-1631]
3	[2-539];[1178-1245]	[1-1];[540-1177]
4	[1-250];[297-383];[386-514];[1025-1064]	[251-296];[384-385];[515-1024];[1065-1623]
5	[27-116];[118-391]	[1-26];[117-117];[392-1454]
6	[1-93];[96-168];[170-262];[264-461]	[94-95];[169-169];[263-263];[462-1639]
7	[1-95];[97-451]	[96-96];[452-1768]
8	[1-502];[1314-1491]	[503-1313];[1492-1510]
9	[1-864]	[865-882]
10	[1-428]	[429-1849]
11	[1-454];[482-514]	[455-481];[515-565]
12	[1-375];[379-511];[533-690];[730-783];[814-1164]	[376-378];[512-532];[691-729];[784-813];[1165-1663]
13	[2-337];[339-556]	[1-1];[338-338];[557-744]
14	[29-366];[368-507]	[1-28];[367-367];[508-1759]
15	[29-366];[368-524]	[1-28];[367-367];[525-1755]
16	[1-641]	[642-936]
17	[1-708];[711-747]	[709-710]
18	[1-639]	[640-1884]
19	[1-631]	[632-691]
20	[3-416];[418-490]	[1-2];[417-417];[491-1138]
21	[1-468]	None
22	[1-720]	None
23	[1-711]	[712-727]
24	[1-469]	[470-470]
25	[1-231];[234-488]	[232-233];[489-987]
26	[1-296];[300-642];[644-737]	[297-299];[643-643];[738-908]
27	[1-306];[308-762]	[307-307]
28	[1-446];[448-1102]	[447-447]
29	[1-436]	None
30	[7-334];[1420-1468];[1474-1614];[1616-1804];[1845-1919]	[1-6];[335-1419];[1469-1473];[1615-1615];[1805-1844];[1920-1938]
31	[1-342];[345-519];[823-893];[977-1016]	[343-344];[520-822];[894-976];[1017-1116]
32	[1-517];[821-891];[975-1014]	[518-820];[892-974];[1015-1114]
33	[36-352];[354-457];[728-832];[834-1096];[1253-1289];[1291-1350];[1352-1412];[1726-1873]	[1-35];[353-353];[458-727];[833-833];[1097-1252];[1290-1290];[1351-1351];[1413-1725];[1874-2072]
34	[1-409]	None
35	[14-105]	[1-13];[106-836]
36	[1-572];[1120-1271]	[573-1119];[1272-1323]
37	[20-98];[100-510];[1591-1681];[1683-1870]	[1-19];[99-99];[511-1590];[1682-1682];[1871-1945]

38	[1-547]	[548-1330]
39	[1-445]	[446-2124]
40	[1-473];[475-528]	[474-474];[529-1159]
41	[16-506];[1587-1866]	[1-15];[507-1586];[1867-1953]
42	[2-234];[244-451];[974-1226]	[1-1];[235-243];[452-973];[1227-1688]
43	[1-455];[1670-1925]	[456-1669];[1926-1942]
44	[1-579];[815-1031]	[580-814];[1032-1657]
45	[1-489];[1012-1264]	[490-1011];[1265-1733]
46	[1-400];[1184-1223];[1225-1705];[1740-1818]	[401-1183];[1224-1224];[1706-1739];[1819-1871]
47	[1-529];[1326-1505]	[530-1325];[1506-1523]
48	[1-131];[133-510];[560-589]	[132-132];[511-559];[590-832]
49	[1-130];[132-509];[559-588]	[131-131];[510-558];[589-831]
50	[1-650];[652-868];[873-913]	[651-651];[869-872];[914-917]
51	[1-504];[515-605]	[505-514];[606-621]
52	[1-535]	[536-673]
53	[2-563]	[1-1];[564-897]
54	[1-527];[802-870];[882-934];[966-1018];[1037-1080]	[528-801];[871-881];[935-965];[1019-1036];[1081-1101]
55	[1-326];[328-505]	[327-327];[506-1047]
56	[1-340]	[341-925]
57	[1-528]	[529-1240]
58	[1-108];[115-151];[154-340];[342-529]	[109-114];[152-153];[341-341];[530-902]
59	[4-485];[1566-1656];[1658-1845]	[1-3];[486-1565];[1657-1657];[1846-1969]
60	[1-283]	[284-1132]
61	[9-468]	[1-8];[469-631]
62	[1-525];[689-722]	[526-688]
63	[1-88];[90-192];[194-265];[296-409]	[89-89];[193-193];[266-295];[410-1442]
64	[1-517]	[518-795]
65	[1-406];[408-739]	[407-407];[740-1236]
66	[1-489];[849-881]	[490-848]
67	[1-505]	[506-524]
68	[1-325];[328-441];[444-504]	[326-327];[442-443];[505-1472]
69	[1-524];[636-715];[717-809];[811-885];[1567-1715]	[525-635];[716-716];[810-810];[886-1566];[1716-1737]
70	[12-487]	[1-11];[488-1637]
71	[12-487]	[1-11];[488-1636]
72	[1-451]	[452-1758]
73	[1-167];[242-464]	[168-241];[465-1647]
74	[1-167];[242-464]	[168-241];[465-1646]
75	[1-471]	[472-1963]
76	[1-358];[360-543];[655-734];[736-828];[830-904];[1586-1734]	[359-359];[544-654];[735-735];[829-829];[905-1585];[1735-1757]
77	[3-34];[36-474];[582-770];[1709-1746];[1748-1785];[1825-1899]	[1-2];[35-35];[475-581];[771-1708];[1747-1747];[1786-1824];[1900-2027]
78	[1-75];[77-319];[914-1052];[1063-1126];[1168-1203]	[76-76];[320-913];[1053-1062];[1127-1167];[1204-1880]

79	[1-425]	[426-584]
80	[1-752];[947-1017];[1084-1170]	[753-946];[1018-1083];[1171-1351]
81	[1-496];[498-720]	[497-497]
82	[1-324]	[325-1029]
83	[1-477];[1474-1529];[1537-1566];[1577-1616];[1622-1662];[1717-1753]	[478-1473];[1530-1536];[1567-1576];[1617-1621];[1663-1716];[1754-1788]
84	[1-496];[499-568];[752-805]	[497-498];[569-751]
85	[1-527]	[528-814]
86	[1-360]	[361-598]
87	[1-78];[80-583];[625-699]	[79-79];[584-624]
88	[1-889]	[890-905]
89	[1-513]	[514-514]
90	[1-122];[124-155];[157-435];[437-517]	[123-123];[156-156];[436-436];[518-518]
91	[1-133];[165-808]	[134-164]
92	[1-725]	[726-737]
93	[1-409]	[410-728]
94	[1-331]	[332-582]
95	[1-410]	[411-1913]
96	[1-501]	[502-670]
97	[1-141];[143-431]	[142-142];[432-939]
98	[1-193]	[194-661]
99	[1-629]	[630-647]
100	[1-520];[862-954];[976-1005]	[521-861];[955-975];[1006-1006]
101	[1-489];[581-961];[1010-1059]	[490-580];[962-1009]
102	[1-485]	[486-514]
103	[1-540]	[541-1158]
104	[1-556]	[557-1563]
105	[1-868];[870-1006]	[869-869];[1007-1621]
106	[1-491]	[492-557]
107	[1-573]	[574-600]
108	[1-457];[586-1110]	[458-585];[1111-1129]
109	[1-521];[655-778]	[522-654]
110	[1-416];[478-614];[616-990];[992-1065];[1068-1283]	[417-477];[615-615];[991-991];[1066-1067];[1284-1301]
111	[1-416];[478-614];[628-989];[991-1064];[1067-1282]	[417-477];[615-627];[990-990];[1065-1066];[1283-1300]
112	[2-429];[1161-1202];[1212-1388];[1392-1589]	[1-1];[430-1160];[1203-1211];[1389-1391];[1590-1617]
113	[1-487]	[488-1634]
114	[1-70];[86-496]	[71-85];[497-693]
115	[1-358];[360-558]	[359-359];[559-784]
116	[1-215];[218-495];[527-607]	[216-217];[496-526];[608-804]
117	[1-466]	[467-484]
118	[1-515];[906-963]	[516-905];[964-985]
119	[1-744];[746-816]	[745-745];[817-839]
120	[1-85];[87-521]	[86-86];[522-583]

121	[1-532]	[533-1024]
122	[1-318];[325-517];[567-660]	[319-324];[518-566];[661-760]
123	[1-498]	[499-594]
124	[1-427]	[428-559]
125	[1-642]	[643-744]
126	[1-341];[350-696]	[342-349];[697-824]
127	[1-482]	[483-526]
128	[1-338]	[339-618]
129	[1-191];[193-429];[450-678]	[192-192];[430-449];[679-776]
130	[19-463];[465-544]	[1-18];[464-464];[545-998]
131	[1-470]	[471-779]
132	[1-533]	[534-1025]
133	[1-498]	[499-607]
134	[1-168];[170-326];[328-471];[552-738]	[169-169];[327-327];[472-551];[739-774]
135	[1-346];[348-395];[440-473]	[347-347];[396-439];[474-611]
136	[1-324];[343-436]	[325-342];[437-925]
137	[1-186];[188-251];[255-517]	[187-187];[252-254];[518-674]
138	[1-488]	[489-1725]
139	[1-101];[103-190];[292-327];[1091-1161];[1228-1314]	[102-102];[191-291];[328-1090];[1162-1227];[1315-1474]
140	[1-465];[516-653]	[466-515]
141	[1-761];[763-857];[912-1326]	[762-762];[858-911];[1327-1490]
142	[1-476]	[477-661]
143	[1-531];[1471-1508];[1510-1547];[1587-1661]	[532-1470];[1509-1509];[1548-1586];[1662-1789]
144	[1-492];[503-536]	[493-502];[537-2006]
145	[1-570]	[571-1096]
146	[1-536];[621-703];[729-1075];[1198-1445]	[537-620];[704-728];[1076-1197];[1446-1666]
147	[1-555];[578-628]	[556-577];[629-1687]
148	[1-444];[1201-1474];[1480-1516]	[445-1200];[1475-1479];[1517-1747]
149	[1-613];[626-658]	[614-625]
150	[4-199];[201-419];[421-492]	[1-3];[200-200];[420-420];[493-2045]
151	[1-509]	[510-788]
152	[1-483];[485-578]	[484-484];[579-1931]
153	[1-497]	[498-514]
154	[5-509];[579-763];[765-1162]	[1-4];[510-578];[764-764];[1163-1183]
155	[1-486];[1095-1500]	[487-1094];[1501-1545]
156	[1-488];[740-797];[799-884];[895-974]	[489-739];[798-798];[885-894];[975-1068]
157	[1-161];[163-565];[567-701]	[162-162];[566-566];[702-1097]
158	[1-496];[692-754]	[497-691];[755-894]
159	[1-483]	[484-703]
160	[1-494]	[495-849]
161	[1-491]	[492-846]
162	[1-505];[575-759];[761-1164]	[506-574];[760-760];[1165-1176]
163	[1-699]	[700-1084]
164	[38-483];[485-556]	[1-37];[484-484];[557-1793]

165	[1-426];[1303-1444];[1717-1755];[1787-1825]	[427-1302];[1445-1716];[1756-1786];[1826-1849]
166	[2-264];[266-446];[448-519]	[1-1];[265-265];[447-447];[520-1748]
167	[1-519];[523-552]	[520-522];[553-1275]
168	[1-457];[466-571]	[458-465];[572-1023]
169	[1-54];[57-501]	[55-56];[502-1085]
170	[1-541]	[542-776]
171	[1-489]	[490-1219]
172	[1-538];[977-1468]	[539-976];[1469-1487]
173	[1-631]	[632-1915]
174	[21-776];[888-967];[969-1061];[1063-1137];[1819-1967]	[1-20];[777-887];[968-968];[1062-1062];[1138-1818];[1968-1990]
175	[1-508]	[509-1971]
176	[1-127];[129-538];[979-1470]	[128-128];[539-978];[1471-1613]
177	[1-535];[973-1173];[1177-1330];[1332-1361]	[536-972];[1174-1176];[1331-1331]
178	[1-599];[626-830];[1082-1113]	[600-625];[831-1081]
179	[1-623];[1377-1406]	[624-1376];[1407-1960]
180	[1-414];[418-464]	[415-417];[465-1443]
181	[1-522];[533-587]	[523-532];[588-605]
182	[1-78];[99-131];[136-327];[1153-1184];[1210-1274];[1284-1319];[1385-1416]	[79-98];[132-135];[328-1152];[1185-1209];[1275-1283];[1320-1384];[1417-1724]
183	[1-512];[617-805];[871-952];[1387-1422];[1621-1661]	[513-616];[806-870];[953-1386];[1423-1620];[1662-1686]
184	[1-453]	[454-463]
185	[1-773]	None
186	[1-413];[423-604];[606-739]	[414-422];[605-605];[740-753]
187	[1-117];[119-401]	[118-118];[402-754]
188	[1-511];[684-870];[872-928];[935-981]	[512-683];[871-871];[929-934];[982-998]
189	[1-605]	None
190	[2-475]	[1-1];[476-526]
191	[1-910]	None
192	[1-101];[103-668]	[102-102]
193	[1-520];[583-637]	[521-582]
194	[1-706]	None
195	[1-145];[150-451];[466-670]	[146-149];[452-465]
196	[1-509]	[510-510]
197	[1-500]	None
198	[1-503];[505-585]	[504-504];[586-667]
199	[1-498]	[499-514]
200	[1-462]	None
201	[1-551]	None
202	[1-482];[484-550]	[483-483]
203	[1-408]	None
204	[1-519];[521-649]	[520-520];[650-665]
205	[1-261];[263-415];[417-640];[642-782]	[262-262];[416-416];[641-641];[783-1008]
206	[1-455]	None

207	[1-402];[410-526]	[403-409];[527-749]
208	[1-520]	[521-594]
209	[1-197];[200-472]	[198-199];[473-2098]
210	[1-311];[314-427]	[312-313];[428-428]
211	[1-689];[735-769]	[690-734]
212	[1-517]	[518-914]
213	[2-576];[756-795];[1390-1441]	[1-1];[577-755];[796-1389];[1442-1489]
214	[1-482]	[483-776]
215	[1-498]	[499-1412]
216	[1-505];[1000-1293];[1295-1408];[1744-1773]	[506-999];[1294-1294];[1409-1743]
217	[1-102];[104-291];[293-467];[486-708];[723-831];[833-900];[910-1031];[1054-1090];[1097-1153]	[103-103];[292-292];[468-485];[709-722];[832-832];[901-909];[1032-1053];[1091-1096];[1154-1251]
218	[1-452]	[453-894]
219	[1-554];[556-598]	[555-555];[599-910]
220	[1-38];[41-95];[98-386];[388-487]	[39-40];[96-97];[387-387];[488-519]
221	[1-34];[38-220];[222-335];[337-518]	[35-37];[221-221];[336-336];[519-632]
222	[1-468]	[469-652]
223	[1-466]	[467-650]
224	[1-466]	[467-502]
225	[1-489];[653-1008]	[490-652];[1009-1739]
226	[1-657]	None
227	[1-480]	[481-888]
228	[1-501]	[502-716]
229	[1-612]	[613-654]
230	[1-477];[485-538]	[478-484];[539-635]
231	[1-476];[484-537]	[477-483];[538-634]
232	[1-367];[371-512]	[368-370];[513-583]
233	[1-305];[307-442];[460-503];[553-646]	[306-306];[443-459];[504-552];[647-753]
234	[1-260];[262-345];[347-454];[473-515];[565-658]	[261-261];[346-346];[455-472];[516-564];[659-762]
235	[1-427]	[428-537]
236	[1-465]	[466-994]
237	[1-471];[496-526];[557-587];[597-637]	[472-495];[527-556];[588-596];[638-662]
238	[1-338];[352-497]	[339-351];[498-1829]
239	[1-501]	[502-1083]
240	[1-515];[1527-1583];[1585-1687];[1692-1831]	[516-1526];[1584-1584];[1688-1691]
241	[1-515];[1526-1582];[1584-1686];[1691-1830]	[516-1525];[1583-1583];[1687-1690]

Table Vb

Seq Id No	Preferentially excluded fragments	Preferentially included fragments
1	[1-540];[556-615];[2061-2096];[2098-2201]	[541-555];[616-2060];[2097-2097]
2	[1-511];[533-619];[621-690];[730-1132]	[512-532];[620-620];[691-729];[1133-1631]
3	[2-539];[1178-1245]	[1-1];[540-1177]
4	[1-250];[297-383];[386-514];[1025-1064]	[251-296];[384-385];[515-1024];[1065-1623]
5	[27-116];[118-391]	[1-26];[117-117];[392-1454]
6	[1-93];[96-168];[170-262];[264-461]	[94-95];[169-169];[263-263];[462-1639]
7	[1-95];[97-451]	[96-96];[452-1768]
8	[1-502];[1314-1491]	[503-1313];[1492-1510]
9	[1-864]	[865-882]
10	[1-428]	[429-1849]
11	[1-454];[482-514]	[455-481];[515-565]
12	[1-375];[379-511];[533-690];[730-783];[814-1164]	[376-378];[512-532];[691-729];[784-813];[1165-1663]
13	[2-337];[339-556]	[1-1];[338-338];[557-744]
14	[29-366];[368-507]	[1-28];[367-367];[508-1759]
15	[29-366];[368-524]	[1-28];[367-367];[525-1755]
16	[1-641]	[642-936]
17	[1-708];[711-747]	[709-710]
18	[1-639]	[640-1884]
19	[1-631]	[632-691]
20	[3-416];[418-490]	[1-2];[417-417];[491-1138]
21	[1-468]	None
22	[1-720]	None
23	[1-711]	[712-727]
24	[1-469]	[470-470]
25	[1-231];[234-488]	[232-233];[489-987]
26	[1-296];[300-642];[644-737]	[297-299];[643-643];[738-908]
27	[1-306];[308-762]	[307-307]
28	[1-446];[448-1102]	[447-447]
29	[1-436]	None
30	[7-334];[1420-1468];[1474-1614];[1616-1804];[1845-1919]	[1-6];[335-1419];[1469-1473];[1615-1615];[1805-1844];[1920-1938]
31	[1-342];[345-519];[823-893];[977-1016]	[343-344];[520-822];[894-976];[1017-1116]
32	[1-517];[821-891];[975-1014]	[518-820];[892-974];[1015-1114]
33	[36-352];[354-457];[728-832];[834-1096];[1253-1289];[1291-1350];[1352-1412];[1726-1873]	[1-35];[353-353];[458-727];[833-833];[1097-1252];[1290-1290];[1351-1351];[1413-1725];[1874-2072]
34	[1-409]	None
35	[14-105]	[1-13];[106-836]
36	[1-572];[1120-1271]	[573-1119];[1272-1323]
37	[20-98];[100-510];[1591-1681];[1683-1870]	[1-19];[99-99];[511-1590];[1682-1682];[1871-1945]
38	[1-547]	[548-1330]

39	[1-445]	[446-2124]
40	[1-473];[475-528]	[474-474];[529-1159]
41	[16-506];[1587-1866]	[1-15];[507-1586];[1867-1953]
42	[2-234];[244-451];[974-1226]	[1-1];[235-243];[452-973];[1227-1688]
43	[1-455];[1670-1925]	[456-1669];[1926-1942]
44	[1-579];[815-1031]	[580-814];[1032-1657]
45	[1-489];[1012-1264]	[490-1011];[1265-1733]
46	[1-400];[1184-1223];[1225-1705];[1740-1818]	[401-1183];[1224-1224];[1706-1739];[1819-1871]
47	[1-529];[1326-1505]	[530-1325];[1506-1523]
48	[1-131];[133-510];[560-589]	[132-132];[511-559];[590-832]
49	[1-130];[132-509];[559-588]	[131-131];[510-558];[589-831]
50	[1-650];[652-868];[873-913]	[651-651];[869-872];[914-917]
51	[1-504];[515-605]	[505-514];[606-621]
52	[1-535]	[536-673]
53	[2-563]	[1-1];[564-897]
54	[1-527];[802-870];[882-934];[966-1018];[1037-1080]	[528-801];[871-881];[935-965];[1019-1036];[1081-1101]
55	[1-326];[328-505]	[327-327];[506-1047]
56	[1-340]	[341-925]
57	[1-528]	[529-1240]
58	[1-108];[115-151];[154-340];[342-529]	[109-114];[152-153];[341-341];[530-902]
59	[4-485];[1566-1656];[1658-1845]	[1-3];[486-1565];[1657-1657];[1846-1969]
60	[1-283]	[284-1132]
61	[9-468]	[1-8];[469-631]
62	[1-525];[689-722]	[526-688]
63	[1-88];[90-192];[194-265];[296-409]	[89-89];[193-193];[266-295];[410-1442]
64	[1-517]	[518-795]
65	[1-406];[408-739]	[407-407];[740-1236]
66	[1-489];[849-881]	[490-848]
67	[1-505]	[506-524]
68	[1-325];[328-441];[444-504]	[326-327];[442-443];[505-1472]
69	[1-524];[636-715];[717-809];[811-885];[1567-1715]	[525-635];[716-716];[810-810];[886-1566];[1716-1737]
70	[12-487]	[1-11];[488-1637]
71	[12-487]	[1-11];[488-1636]
72	[1-451]	[452-1758]
73	[1-167];[242-464]	[168-241];[465-1647]
74	[1-167];[242-464]	[168-241];[465-1646]
75	[1-471]	[472-1963]
76	[1-358];[360-543];[655-734];[736-828];[830-904];[1586-1734]	[359-359];[544-654];[735-735];[829-829];[905-1585];[1735-1757]
77	[3-34];[36-474];[582-770];[1709-1746];[1748-1785];[1825-1899]	[1-2];[35-35];[475-581];[771-1708];[1747-1747];[1786-1824];[1900-2027]
78	[1-75];[77-319];[914-1052];[1063-1126];[1168-1203]	[76-76];[320-913];[1053-1062];[1127-1167];[1204-1880]
79	[1-425]	[426-584]

80	[1-752];[947-1017];[1084-1170]	[753-946];[1018-1083];[1171-1351]
81	[1-496];[498-720]	[497-497]
82	[1-324]	[325-1029]
83	[1-477];[1474-1529];[1537-1566];[1577-1616];[1622-1662];[1717-1753]	[478-1473];[1530-1536];[1567-1576];[1617-1621];[1663-1716];[1754-1788]
84	[1-496];[499-568];[752-805]	[497-498];[569-751]
85	[1-527]	[528-814]
86	[1-360]	[361-598]
87	[1-78];[80-583];[625-699]	[79-79];[584-624]
88	[1-889]	[890-905]
89	[1-513]	[514-514]
90	[1-122];[124-155];[157-435];[437-517]	[123-123];[156-156];[436-436];[518-518]
91	[1-133];[165-808]	[134-164]
92	[1-725]	[726-737]
93	[1-409]	[410-728]
94	[1-331]	[332-582]
95	[1-410]	[411-1913]
96	[1-501]	[502-670]
97	[1-141];[143-431]	[142-142];[432-939]
98	[1-193]	[194-661]
99	[1-629]	[630-647]
100	[1-520];[862-954];[976-1005]	[521-861];[955-975];[1006-1006]
101	[1-489];[581-961];[1010-1059]	[490-580];[962-1009]
102	[1-485]	[486-514]
103	[1-540]	[541-1158]
104	[1-556]	[557-1563]
105	[1-868];[870-1006]	[869-869];[1007-1621]
106	[1-491]	[492-557]
107	[1-573]	[574-600]
108	[1-457];[586-1110]	[458-585];[1111-1129]
109	[1-521];[655-778]	[522-654]
110	[1-416];[478-614];[616-990];[992-1065];[1068-1283]	[417-477];[615-615];[991-991];[1066-1067];[1284-1301]
111	[1-416];[478-614];[628-989];[991-1064];[1067-1282]	[417-477];[615-627];[990-990];[1065-1066];[1283-1300]
112	[2-429];[1161-1202];[1212-1388];[1392-1589]	[1-1];[430-1160];[1203-1211];[1389-1391];[1590-1617]
113	[1-487]	[488-1634]
114	[1-70];[86-496]	[71-85];[497-693]
115	[1-358];[360-558]	[359-359];[559-784]
116	[1-215];[218-495];[527-607]	[216-217];[496-526];[608-804]
117	[1-466]	[467-484]
118	[1-515];[906-963]	[516-905];[964-985]
119	[1-744];[746-816]	[745-745];[817-839]
120	[1-85];[87-521]	[86-86];[522-583]
121	[1-532]	[533-1024]

122	[1-318];[325-517];[567-660]	[319-324];[518-566];[661-760]
123	[1-498]	[499-594]
124	[1-427]	[428-559]
125	[1-642]	[643-744]
126	[1-341];[350-696]	[342-349];[697-824]
127	[1-482]	[483-526]
128	[1-338]	[339-618]
129	[1-191];[193-429];[450-678]	[192-192];[430-449];[679-776]
130	[19-463];[465-544]	[1-18];[464-464];[545-998]
131	[1-470]	[471-779]
132	[1-533]	[534-1025]
133	[1-498]	[499-607]
134	[1-168];[170-326];[328-471];[552-738]	[169-169];[327-327];[472-551];[739-774]
135	[1-346];[348-395];[440-473]	[347-347];[396-439];[474-611]
136	[1-324];[343-436]	[325-342];[437-925]
137	[1-186];[188-251];[255-517]	[187-187];[252-254];[518-674]
138	[1-488]	[489-1725]
139	[1-101];[103-190];[292-327];[1091-1161];[1228-1314]	[102-102];[191-291];[328-1090];[1162-1227];[1315-1474]
140	[1-465];[516-653]	[466-515]
141	[1-761];[763-857];[912-1326]	[762-762];[858-911];[1327-1490]
142	[1-476]	[477-661]
143	[1-531];[1471-1508];[1510-1547];[1587-1661]	[532-1470];[1509-1509];[1548-1586];[1662-1789]
144	[1-492];[503-536]	[493-502];[537-2006]
145	[1-570]	[571-1096]
146	[1-536];[621-703];[729-1075];[1198-1445]	[537-620];[704-728];[1076-1197];[1446-1666]
147	[1-555];[578-628]	[556-577];[629-1687]
148	[1-444];[1201-1474];[1480-1516]	[445-1200];[1475-1479];[1517-1747]
149	[1-613];[626-658]	[614-625]
150	[4-199];[201-419];[421-492]	[1-3];[200-200];[420-420];[493-2045]
151	[1-509]	[510-788]
152	[1-483];[485-578]	[484-484];[579-1931]
153	[1-497]	[498-514]
154	[5-509];[579-763];[765-1162]	[1-4];[510-578];[764-764];[1163-1183]
155	[1-486];[1095-1500]	[487-1094];[1501-1545]
156	[1-488];[740-797];[799-884];[895-974]	[489-739];[798-798];[885-894];[975-1068]
157	[1-161];[163-565];[567-701]	[162-162];[566-566];[702-1097]
158	[1-496];[692-754]	[497-691];[755-894]
159	[1-483]	[484-703]
160	[1-494]	[495-849]
161	[1-491]	[492-846]
162	[1-505];[575-759];[761-1164]	[506-574];[760-760];[1165-1176]
163	[1-699]	[700-1084]
164	[38-483];[485-556]	[1-37];[484-484];[557-1793]
165	[1-426];[1303-1444];[1717-1755];[1787-1825]	[427-1302];[1445-1716];[1756-1786];[1826-1849]

166	[2-264];[266-446];[448-519]	[1-1];[265-265];[447-447];[520-1748]
167	[1-519];[523-552]	[520-522];[553-1275]
168	[1-457];[466-571]	[458-465];[572-1023]
169	[1-54];[57-501]	[55-56];[502-1085]
170	[1-541]	[542-776]
171	[1-489]	[490-1219]
172	[1-538];[977-1468]	[539-976];[1469-1487]
173	[1-631]	[632-1915]
174	[21-776];[888-967];[969-1061];[1063-1137];[1819-1967]	[1-20];[777-887];[968-968];[1062-1062];[1138-1818];[1968-1990]
175	[1-508]	[509-1971]
176	[1-127];[129-538];[979-1470]	[128-128];[539-978];[1471-1613]
177	[1-535];[973-1173];[1177-1330];[1332-1361]	[536-972];[1174-1176];[1331-1331]
178	[1-599];[626-830];[1082-1113]	[600-625];[831-1081]
179	[1-623];[1377-1406]	[624-1376];[1407-1960]
180	[1-414];[418-464]	[415-417];[465-1443]
181	[1-522];[533-587]	[523-532];[588-605]
182	[1-78];[99-131];[136-327];[1153-1184];[1210-1274];[1284-1319];[1385-1416]	[79-98];[132-135];[328-1152];[1185-1209];[1275-1283];[1320-1384];[1417-1724]
183	[1-512];[617-805];[871-952];[1387-1422];[1621-1661]	[513-616];[806-870];[953-1386];[1423-1620];[1662-1686]
184	[1-453]	[454-463]
185	[1-773]	None
186	[1-413];[423-604];[606-739]	[414-422];[605-605];[740-753]
187	[1-117];[119-401]	[118-118];[402-754]
188	[1-511];[684-870];[872-928];[935-981]	[512-683];[871-871];[929-934];[982-998]
189	[1-605]	None
190	[2-475]	[1-1];[476-526]
191	[1-910]	None
192	[1-101];[103-668]	[102-102]
193	[1-520];[583-637]	[521-582]
194	[1-706]	None
195	[1-145];[150-451];[466-670]	[146-149];[452-465]
196	[1-509]	[510-510]
197	[1-500]	None
198	[1-503];[505-585]	[504-504];[586-667]
199	[1-498]	[499-514]
200	[1-462]	None
201	[1-551]	None
202	[1-482];[484-550]	[483-483]
203	[1-408]	None
204	[1-519];[521-649]	[520-520];[650-665]
205	[1-261];[263-415];[417-640];[642-782]	[262-262];[416-416];[641-641];[783-1008]
206	[1-455]	None
207	[1-402];[410-526]	[403-409];[527-749]
208	[1-520]	[521-594]

209	[1-197];[200-472]	[198-199];[473-2098]
210	[1-311];[314-427]	[312-313];[428-428]
211	[1-689];[735-769]	[690-734]
212	[1-517]	[518-914]
213	[2-576];[756-795];[1390-1441]	[1-1];[577-755];[796-1389];[1442-1489]
214	[1-482]	[483-776]
215	[1-498]	[499-1412]
216	[1-505];[1000-1293];[1295-1408];[1744-1773]	[506-999];[1294-1294];[1409-1743]
217	[1-102];[104-291];[293-467];[486-708];[723-831];[833-900];[910-1031];[1054-1090];[1097-1153]	[103-103];[292-292];[468-485];[709-722];[832-832];[901-909];[1032-1053];[1091-1096];[1154-1251]
218	[1-452]	[453-894]
219	[1-554];[556-598]	[555-555];[599-910]
220	[1-38];[41-95];[98-386];[388-487]	[39-40];[96-97];[387-387];[488-519]
221	[1-34];[38-220];[222-335];[337-518]	[35-37];[221-221];[336-336];[519-632]
222	[1-468]	[469-652]
223	[1-466]	[467-650]
224	[1-466]	[467-502]
225	[1-489];[653-1008]	[490-652];[1009-1739]
226	[1-657]	None
227	[1-480]	[481-888]
228	[1-501]	[502-716]
229	[1-612]	[613-654]
230	[1-477];[485-538]	[478-484];[539-635]
231	[1-476];[484-537]	[477-483];[538-634]
232	[1-367];[371-512]	[368-370];[513-583]
233	[1-305];[307-442];[460-503];[553-646]	[306-306];[443-459];[504-552];[647-753]
234	[1-260];[262-345];[347-454];[473-515];[565-658]	[261-261];[346-346];[455-472];[516-564];[659-762]
235	[1-427]	[428-537]
236	[1-465]	[466-994]
237	[1-471];[496-526];[557-587];[597-637]	[472-495];[527-556];[588-596];[638-662]
238	[1-338];[352-497]	[339-351];[498-1829]
239	[1-501]	[502-1083]
240	[1-515];[1527-1583];[1585-1687];[1692-1831]	[516-1526];[1584-1584];[1688-1691]
241	[1-515];[1526-1582];[1584-1686];[1691-1830]	[516-1525];[1583-1583];[1687-1690]

Table VI

Seq Id No	Designation of domain	Database	Positions of domains
242	Cell attachment sequence	PROSITE	141-143
242	Peptidase family M20/M25/M40	PFAM	107-451
244	Mitochondrial energy transfer proteins signature	PROSITE	26-35
244	Mitochondrial energy transfer proteins signature	PROSITE	199-208
244	Mitochondrial carrier proteins	PFAM	5-84;87-175;178-272
244	Mitochondrial energy transfer proteins.	BLOCKSPLUS	12-36
244	Mitochondrial energy transfer proteins.	BLOCKSPLUS	13-36
244	Mitochondrial energy transfer proteins.	BLOCKSPLUS	131-144
245	Leucine zipper pattern	PROSITE	371-392
249	Leucine zipper pattern	PROSITE	20-41
251	Mitochondrial energy transfer proteins signature	PROSITE	26-35
251	Mitochondrial carrier proteins	PFAM	5-72
251	Mitochondrial energy transfer proteins.	BLOCKSPLUS	12-36
251	Mitochondrial energy transfer proteins.	BLOCKSPLUS	13-36
254	Pancreatic ribonuclease family signature	PROSITE	63-69
254	Pancreatic ribonucleases	PFAM	26-143
254	PANCREATIC RIBONUCLEASE FAMILY SIGNATURE	BLOCKSPLUS	49-69
254	Pancreatic ribonuclease family proteins.	BLOCKSPLUS	115-140
254	PANCREATIC RIBONUCLEASE FAMILY SIGNATURE	BLOCKSPLUS	92-110
254	PANCREATIC RIBONUCLEASE FAMILY SIGNATURE	BLOCKSPLUS	114-133
254	Pancreatic ribonuclease family proteins.	BLOCKSPLUS	30-40
254	PANCREATIC RIBONUCLEASE FAMILY SIGNATURE	BLOCKSPLUS	114-137
254	PANCREATIC RIBONUCLEASE FAMILY SIGNATURE	BLOCKSPLUS	69-86
255	L-lactate dehydrogenase active site	PROSITE	239-245
255	lactate/malate dehydrogenase	PFAM	71-380
255	L-lactate dehydrogenase proteins.	BLOCKSPLUS	186-224
255	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	96-121
255	L-lactate dehydrogenase proteins.	BLOCKSPLUS	71-102
255	L-lactate dehydrogenase proteins.	BLOCKSPLUS	238-256
255	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	183-203
255	L-lactate dehydrogenase proteins.	BLOCKSPLUS	288-323

255	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	207-224
255	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	71-92
255	L-lactate dehydrogenase proteins.	BLOCKSPLUS	138-167
256	lactate/malate dehydrogenase	PFAM	71-124
256	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	96-121
256	L-lactate dehydrogenase proteins.	BLOCKSPLUS	71-102
256	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	71-92
256	L-lactate dehydrogenase proteins.	BLOCKSPLUS	71-100
256	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	71-84
257	Leucine zipper pattern	PROSITE	155-176
259	HORMA domain	PFAM	22-230
261	Leucine zipper pattern	PROSITE	142-163
261	Leucine zipper pattern	PROSITE	170-191
263	Leucine zipper pattern	PROSITE	15-36
264	Ubiquitin family	PFAM	1-82
264	Ubiquitin domain proteins.	BLOCKSPLUS	17-62
264	Ubiquitin domain proteins.	BLOCKSPLUS	21-68
264	Ubiquitin domain proteins.	BLOCKSPLUS	26-68
264	Ubiquitin domain proteins.	BLOCKSPLUS	17-68
266	u-PAR/Ly-6 domain	PFAM	60-119
266	Squash family of serine protease inhibit	PFAM	32-47
267	Zinc finger, C2H2 type, domain proteins.	BLOCKSPLUS	185-202
271	LBP / BPI / CETP family signature	PROSITE	28-60
271	Pyrokinins signature	PROSITE	324-328
271	LBP / BPI / CETP family	PFAM	10-479
271	LBP / BPI / CETP family proteins.	BLOCKSPLUS	72-118
271	LBP / BPI / CETP family proteins.	BLOCKSPLUS	209-253
271	LBP / BPI / CETP family proteins.	BLOCKSPLUS	28-58
271	LBP / BPI / CETP family proteins.	BLOCKSPLUS	275-309
271	LBP / BPI / CETP family proteins.	BLOCKSPLUS	76-113
272	Zinc finger, C3HC4 type (RING finger), signature	PROSITE	102-111
272	Zinc finger, C3HC4 type (RING finger)	PFAM	87-129
272	Zinc finger, C3HC4 type (RING finger), proteins.	BLOCKSPLUS	102-111
273	Zinc finger, C3HC4 type (RING finger), signature	PROSITE	30-39
273	Zinc finger, C3HC4 type (RING finger)	PFAM	15-57
273	Zinc finger, C3HC4 type (RING finger), proteins.	BLOCKSPLUS	30-39
274	RNA 3'-terminal phosphate cyclase signature	PROSITE	157-167
274	RNA 3'-terminal phosphate cyclase	PFAM	1-368

274	RNA 3'-terminal phosphate cyclase proteins.	BLOCKSPLUS	12-44
274	RNA 3'-terminal phosphate cyclase proteins.	BLOCKSPLUS	157-168
275	Ribosomal L27 protein	PFAM	31-86
277	Cell attachment sequence	PROSITE	292-294
277	DHHC zinc finger domain	PFAM	140-204
279	Endogenous opioids neuropeptides precursors signature	PROSITE	26-65
279	Vertebrate endogenous opioids neurope	PFAM	3-257
279	Endogenous opioids neuropeptides precursors proteins.	BLOCKSPLUS	100-126
279	Endogenous opioids neuropeptides precursors proteins.	BLOCKSPLUS	209-237
279	Endogenous opioids neuropeptides precursors proteins.	BLOCKSPLUS	43-66
279	Endogenous opioids neuropeptides precursors proteins.	BLOCKSPLUS	18-38
279	Endogenous opioids neuropeptides precursors proteins.	BLOCKSPLUS	24-36
279	Endogenous opioids neuropeptides precursors proteins.	BLOCKSPLUS	105-125
280	Leucine zipper pattern	PROSITE	136-157
280	Leucine zipper pattern	PROSITE	272-293
283	Immunoglobulins and major histocompatibility complex proteins signature	PROSITE	380-386
283	Immunoglobulin domain	PFAM	205-285;318-384
283	Immunoglobulins and major histocompatibility complex proteins.	BLOCKSPLUS	319-336
284	Fucosyl transferase	PFAM	70-406
285	FAD/NAD-binding Cytochrome reductase	PFAM	27-149
285	Oxidoreductase FAD/NAD-binding domain	PFAM	176-290
285	Eukaryotic molybdopterin oxidoreductases proteins.	BLOCKSPLUS	58-86
285	CYTOCHROME B5 REDUCTASE SIGNATURE	BLOCKSPLUS	75-86
285	CYTOCHROME B5 REDUCTASE SIGNATURE	BLOCKSPLUS	274-283
285	CYTOCHROME B5 REDUCTASE SIGNATURE	BLOCKSPLUS	141-156
285	Eukaryotic molybdopterin oxidoreductases proteins.	BLOCKSPLUS	274-286
285	Eukaryotic molybdopterin oxidoreductases proteins.	BLOCKSPLUS	60-85
285	CYTOCHROME B5 REDUCTASE SIGNATURE	BLOCKSPLUS	181-198
285	FLAVOPROTEIN PYRIDINE NUCLEOTIDE CYTOCHROME REDUCTASE SIGNATURE	BLOCKSPLUS	181-197

286	Immunoglobulins and major histocompatibility complex proteins signature	PROSITE	380-386
286	Immunoglobulin domain	PFAM	205-285;318-384
286	Immunoglobulins and major histocompatibility complex proteins.	BLOCKSPLUS	319-336
287	Leucine zipper pattern	PROSITE	126-147
288	Leucine zipper pattern	PROSITE	20-41
291	Tissue inhibitors of metalloproteinases signature	PROSITE	24-36
291	Tissue inhibitor of metalloproteinases	PFAM	22-199
291	Tissue inhibitors of metalloproteinases proteins.	BLOCKSPLUS	21-46
291	Tissue inhibitors of metalloproteinases proteins.	BLOCKSPLUS	106-148
291	Tissue inhibitors of metalloproteinases proteins.	BLOCKSPLUS	81-95
291	Tissue inhibitors of metalloproteinases proteins.	BLOCKSPLUS	61-72
294	Domain of unknown function DUF59	PFAM	31-135
296	Immunoglobulin domain	PFAM	141-197
297	TonB-dependent receptor proteins signature 1	PROSITE	1-42
298	Fibroblast growth factor	PFAM	48-129
299	BolA-like protein	PFAM	39-114
299	PROTEIN BOLA TRANSCRIPTION REGULATION AC.	BLOCKSPLUS	68-98
301	Cell attachment sequence	PROSITE	172-174
303	Ribosomal L27 protein	PFAM	31-115
304	Leucine rich repeat C-terminal domain	PFAM	173-222
304	Leucine Rich Repeat	PFAM	92-115;116-139;140-163;164-185
309	Leucine rich repeat C-terminal domain	PFAM	173-222
309	Leucine Rich Repeat	PFAM	92-115;116-139;140-163;164-185
311	NOL1/NOP2/sun family	PFAM	201-276;353-378
311	NOL1/NOP2/sun family proteins.	BLOCKSPLUS	230-245
311	NOL1/NOP2/sun family proteins.	BLOCKSPLUS	231-245
312	NOL1/NOP2/sun family	PFAM	201-276
312	NOL1/NOP2/sun family proteins.	BLOCKSPLUS	230-245
312	NOL1/NOP2/sun family proteins.	BLOCKSPLUS	231-245
314	Leucine zipper pattern	PROSITE	8-29
315	Leucine zipper pattern	PROSITE	8-29
341	Immunoglobulin domain	PFAM	45-112
349	CDP-alcohol phosphatidyltransferases signature	PROSITE	54-76
349	Cytochrome b/b6 Qo site signature	PROSITE	97-102
354	SAM domain (Sterile alpha motif)	PFAM	82-147
361	Ribosomal Proteins L2	PFAM	96-124
368	DAD family	PFAM	1-78
370	Ribosomal protein L34	PFAM	51-92

385	Kelch motif	PFAM	20-66;68-114;116-162;164-209;211-265;270-316
386	SPRY domain	PFAM	85-205
388	PHD-finger.	BLOCKSPLUS	329-339
389	Eukaryotic thiol (cysteine) proteases histidine active site	PROSITE	268-278
389	Heat shock hsp70 proteins family signature 3	PROSITE	332-346
389	Hsp70 protein	PFAM	3-509
390	Eukaryotic-type carbonic anhydrase	PFAM	20-59
391	PMP-22/EMP/MP20/Claudin family	PFAM	4-162
392	Sec1 family.	BLOCKSPLUS	89-107
393	PMP-22/EMP/MP20/Claudin family	PFAM	4-182
394	Myc-type, 'helix-loop-helix' dimerization domain signature	PROSITE	13-28
395	Glutathione S-transferases.	PFAM	47-122;260-309
396	Transmembrane 4 family signature	PROSITE	112-134
396	Transmembrane 4 family	PFAM	66-273
396	Transmembrane 4 family proteins.	BLOCKSPLUS	108-146
396	TRANSMEMBRANE FOUR FAMILY SIGNATURE	BLOCKSPLUS	129-151
396	TRANSMEMBRANE FOUR FAMILY SIGNATURE	BLOCKSPLUS	108-127
396	TRANSMEMBRANE FOUR FAMILY SIGNATURE	BLOCKSPLUS	247-274
396	TRANSMEMBRANE FOUR FAMILY SIGNATURE	BLOCKSPLUS	129-150
396	TRANSMEMBRANE FOUR FAMILY SIGNATURE	BLOCKSPLUS	128-154
397	ATP/GTP-binding site motif A (P-loop)	PROSITE	6-13
397	ADP-ribosylation factor family	PFAM	2-172
398	Isochorismatase family	PFAM	17-147
399	PAP2 superfamily	PFAM	19-175
400	Zinc carboxypeptidases, zinc-binding region 2 signature	PROSITE	117-127
401	Zinc finger, C2H2 type, domain	PROSITE	36-57
401	Zinc finger, C2H2 type, domain	PROSITE	73-93
401	Zinc finger, C2H2 type, domain	PROSITE	114-134
401	Zinc finger, C2H2 type, domain	PROSITE	145-165
401	Zinc finger, C2H2 type	PFAM	34-57;71-93;112-134;143-165
401	Zinc finger, C2H2 type, domain proteins.	BLOCKSPLUS	145-162
401	Zinc finger, C2H2 type, domain proteins.	BLOCKSPLUS	114-131
401	Zinc finger, C2H2 type, domain proteins.	BLOCKSPLUS	73-90
402	Zinc finger, C2H2 type, domain	PROSITE	113-133

402	Zinc finger, C2H2 type, domain	PROSITE	144-164
402	Regulator of chromosome condensation (RCC1) signature 2	PROSITE	65-75
402	Zinc finger, C2H2 type	PFAM	111-133;142-164
402	Zinc finger, C2H2 type, domain proteins.	BLOCKSPLUS	144-161
402	Zinc finger, C2H2 type, domain proteins.	BLOCKSPLUS	113-130
403	Glutathione S-transferases.	PFAM	47-122;260-309
405	PMP-22/EMP/MP20/Claudin family	PFAM	4-182
406	WD domain, G-beta repeat	PFAM	267-304;333-370
408	Rhomboid family	PFAM	186-323
410	Ank repeat	PFAM	47-79
410	REPEAT PROTEIN ANK NUCLEAR ANKYR.	BLOCKSPLUS	78-89
410	Ank repeat proteins.	BLOCKSPLUS	48-56
412	Serine proteases, subtilase family, aspartic acid proteins.	BLOCKSPLUS	165-178
414	Sir2 family	PFAM	84-268
416	Kelch motif	PFAM	20-66;68-114;116-162;164-209;211-265;270-316
418	Zinc-binding dehydrogenases	PFAM	16-313
426	Leucine zipper pattern	PROSITE	144-165
447	Cytochrome c family heme-binding site signature	PROSITE	19-24
447	Immunoglobulins and major histocompatibility complex proteins signature	PROSITE	17-23
453	eIF-6 family	PFAM	3-103
454	Cell attachment sequence	PROSITE	226-228
456	Leucine zipper pattern	PROSITE	211-232
457	Leucine zipper pattern	PROSITE	236-257
466	Zinc finger, C3HC4 type (RING finger), signature	PROSITE	56-65
466	SPRY domain	PFAM	375-500
466	Zinc finger, C3HC4 type (RING finger)	PFAM	41-81
466	B-box zinc finger.	PFAM	110-153
466	Domain in SPla and the RYanodine Receptor.	BLOCKSPLUS	359-381
466	Domain in SPla and the RYanodine Receptor.	BLOCKSPLUS	443-457
466	Domain in SPla and the RYanodine Receptor.	BLOCKSPLUS	359-380
466	Zinc finger, C3HC4 type (RING finger), proteins.	BLOCKSPLUS	56-65
479	UBX domain	PFAM	329-408
481	TBC domain	PFAM	65-171
481	Probable rabGAP domain proteins.	BLOCKSPLUS	153-159
482	TBC domain	PFAM	65-177

482	Probable rabGAP domain proteins.	BLOCKSPLUS	153-159

Table VII

Seq Id No	Epitopes
242	98..109;119..127;136..147;156..170;242..248;255..265;318..328;356..363;399..407;443..450;475..490
243	3..9;59..65;69..79;113..126;142..155;193..198;212..220;231..245;302..315
244	29..36;33..42;79..87;139..147;269..274
245	101..107;141..151;156..165;196..207;225..233;242..251;253..260;284..298;323..330;339..347;395..406
247	41..51;108..120;121..131;190..200;255..261;302..307
248	5..11;38..46;52..60;75..83;92..99;133..150;167..183;187..200;210..219;244..252;270..286;335..345;354..371;390..397
249	68..80;91..99;132..138;185..193;265..273;276..293;295..306;305..329;327..341;347..358;394..403
250	28..37;60..67;73..81
251	33..45;64..71
252	20..30;35..45;49..59;74..83
253	3..9;59..65
254	22..33;35..52;53..67;70..77;80..100;106..117;142..147
255	116..123;147..156;201..208;262..278
256	10..15;116..121
257	41..51;52..66;72..80;94..101;120..127;134..147;180..193;204..210;227..240
258	147..157;189..199
259	52..59;66..76;103..113;115..127;131..140;143..148;181..199;242..250;253..262;262..273;279..289;330..341;342..366;373..394
260	94..107;112..119;125..134
261	121..126;144..152;213..224
263	44..50
264	51..58;82..90;153..164
266	15..20;38..49;76..81;95..105
267	74..91;94..99;117..130;140..154;153..161;175..184;201..210;228..240;250..255
268	36..42;43..54
269	41..46;64..73;80..100;106..122;160..172
270	38..48;82..88
271	34..40;72..79;111..123;146..153;251..259;307..314;316..322;372..377;436..444
272	12..17;51..58;75..85;128..136
273	4..13;56..64
274	34..46;120..127;157..163;182..191;231..240;259..267;273..279;291..299;344..355
275	30..55;72..78
276	27..35;37..45;49..61;61..77;102..109;144..152;170..180;179..188
277	61..67;147..152;154..166;284..299;308..313
278	72..82;451..461;532..541
279	24..31;72..84;83..92;97..111;144..149;161..182;181..189;192..198;204..214;216..233;241..254;256..263

280	15..20;50..56;61..66;204..212;251..260;354..362
281	24..38;44..52
282	72..82;451..461;532..541
283	1..6;21..31;77..83;115..120;228..237;276..281;335..343;401..407;440..456;456..468
284	4..9;39..47;50..66;78..94;111..122;132..141;169..174;190..202;213..220;243..252;261..274;282..300;369..376;379..389;395..403
285	29..38;42..47;58..65;100..110;121..134;156..161;161..173;201..207;230..239;243..254;290..302
286	1..6;21..28;77..83;115..120;228..237;276..281;335..343;401..407
287	2..10;94..104;248..258;268..286
288	68..80;91..99;132..138;185..194;262..270;273..288;291..301;300..324;322..336;342..353;389..398
289	23..38
291	28..35;96..104;134..144;159..167;177..187;191..198
292	1..7;56..64;66..73;77..92
293	40..45;99..109
294	47..57;120..126
295	31..61;76..82;143..149;156..169
296	133..143;151..156;161..167;169..181;185..194
297	50..58;59..69;113..123;120..137
298	45..55;52..63;106..117;118..128;126..131;148..155;157..164;172..190;212..221;232..247
299	51..59;82..87;113..125;124..135
300	72..82;451..461;532..541
301	43..52;88..105;192..211;255..271
302	3..18;37..44;57..65;70..76;98..113;121..134
303	30..55;72..77;82..88;106..113
304	2..11;33..42;48..54;55..63;122..131;147..154;168..180;200..209;211..220;226..233;268..278;286..291
305	22..31
306	5..11;25..35;72..81;124..134;147..157;163..178;177..186;185..195;207..217
307	23..38
308	66..72;84..100
309	2..11;33..42;48..54;55..63;122..131;147..154;168..180;200..209;211..220;226..233;268..278;286..291
310	45..52;60..68;88..94;99..109;113..120;121..134;162..171;169..184;194..202;209..215;223..235;239..248;273..281;292..301;319..329;336..341;389..394;398..405;421..426
311	15..21;28..35;82..91;113..120;125..133;153..167;236..243;291..298;307..312;316..327;390..396;406..413;436..457
312	15..21;28..35;82..91;113..120;125..133;153..167;236..243;291..298;307..312;316..327;352..370;370..382
313	38..46;52..60;75..83;92..99;133..150;167..183;187..200;210..219;239..256
314	36..42;52..58;65..70;80..87;143..155;161..168;176..185;203..208;263..272
315	36..42;52..58;65..70;80..87;143..155;161..168
316	33..47;49..58;106..117;125..132

317	45..53;54..68;88..94;99..109;113..120;121..134;162..171;169..184;194..202;209..215;223..235;239..248;273..281;292..301;319..329;336..341;389..394;398..405;421..426
318	42..54;69..77;124..130;148..153;156..165;186..210;219..227;271..286;293..300
319	15..21;25..37;36..59;58..64;80..89;86..102;105..119
320	1..6;81..89
321	111..116
322	1..21;50..68;76..85
323	11..16;49..68
325	14..20;40..55;69..76;122..131;128..138;158..166
326	18..45;61..68;81..89;110..141;142..149
327	43..48;53..62;85..95
328	4..9;34..46
329	1..7;58..69
330	34..42;46..52;56..66
331	59..69;72..80;80..89;90..98;110..115
332	14..24;50..56
333	1..6;41..47;86..96;120..134
334	17..23;41..47;50..57;85..90;96..105;154..159;181..192;192..198
335	4..9;21..28
336	49..65;70..80
337	20..30;36..42;53..61;74..83;110..119;125..138;137..142
338	21..55;55..65;62..81;88..100;99..107
339	33..47;52..60;73..82
340	5..12;13..21;32..51;65..72
341	19..30;44..52;51..61;68..82;96..108
342	18..26;44..54;71..77
343	21..33;57..66;68..77;80..89;91..98
344	36..44;89..95;110..116;181..193
345	19..28;34..41;43..55;80..86;119..124;159..165;167..176
346	26..35;39..49;88..95;136..145;207..217
347	2..10;26..32;52..68;75..86
348	71..77;82..89;114..125
349	54..66;70..76;294..302
350	16..25;31..37;69..80
351	68..73;118..129;135..148
352	68..73;118..129;135..148
353	10..18;61..69;68..77
354	13..20;44..51;57..85;94..102;143..151
355	18..30;38..43;43..54
356	1..6;40..45;120..126;129..154;153..162
357	49..59;133..141;152..167
358	9..26;80..85;96..102
359	98..104;181..187
360	1..6;44..54;113..123;147..161
361	64..74;77..90;112..141
362	20..30;37..42;53..61;74..83;110..119;125..134;138..147
363	1..11;26..31;53..60;97..105;110..117;141..146
364	16..24
365	7..15

367	8..14
368	16..23
369	5..14;59..71;74..83
370	39..45;45..67;82..91
371	37..46;81..86;92..99;140..156;163..169;179..184;209..216;242..252;258..268
372	72..81;93..104;141..146
373	71..77;86..93
374	1..7;31..47;47..57;59..65
375	5..10;55..60;83..90
376	34..43;40..46;67..81
377	27..38;49..56;54..64;68..78;111..122
378	9..26;80..85;96..102
379	41..50;52..58;195..204;318..327;337..347
380	98..103
381	46..51
382	67..72
383	4..12;25..33;73..81
384	42..54;69..77;117..127;125..141
385	105..112;146..155;181..187;200..214;230..243;254..264;261..272;290..295
386	46..53;92..100;107..114;116..123;126..132;176..182
387	5..12;44..50;158..167;190..195
388	16..25;31..37;73..79;96..101;105..110;158..166;215..224;240..247;273..280;299..306;321..334;347..352;370..375
389	28..35;72..81;93..100;101..113;183..192;218..228;284..300;318..330;339..344;443..454;481..489;491..500
390	1..8;39..48
391	104..114;157..162
392	28..34;31..36;46..55;77..85
393	104..114;188..224
394	44..51;107..114
395	2..15;51..61;82..88;104..114;249..259;286..298;333..340;361..367
396	24..31;54..65;79..85;92..99;180..186;216..221
397	20..33;31..41;67..75;82..89;168..173
398	58..65;93..101;135..143;198..203
399	11..17;37..48;71..79;94..100;99..112;132..144;161..173;173..180
400	21..31;65..84;91..99
401	1..9;11..27;78..83;88..96;107..112;112..121;135..141;147..153;164..170
402	1..9;11..35;83..94;106..111;111..120;134..140;146..152;163..169
403	2..15;51..61;82..88;104..114;249..259;286..298;333..340;361..367
405	104..114;188..224
406	1..7;61..71;77..83;80..85;163..179;204..211;210..225;231..238;254..259;272..278;305..321;347..356
408	53..77;120..130;144..159;159..169;196..202;266..272;331..345
409	1..9;73..81;226..236

410	12..21;37..49;78..85
411	22..36;151..156;161..170
412	56..65;136..146;179..186;192..207;227..236
413	8..14;92..99;184..191
414	1..10;16..30;71..77;95..105;102..113;123..130;137..145;165..171;199..205;217..226;286..295;309..314;360..376;384..389
415	19..34;100..107;115..123;143..149;154..164;168..175;176..189;217..226;224..239;249..257;264..270;278..290;294..303;328..336;347..356;374..384;391..396;444..449;453..460;476..481
416	105..112;147..155;181..187;200..214;230..243;254..264;261..272;290..295
417	8..14
418	8..14;92..99;227..236;260..273;293..300
419	8..16;117..122
420	27..35;71..83;91..97;137..146
421	2..11;77..82
422	20..26;61..69
423	9..19;25..34;47..60;57..65;87..92;106..116;126..134
424	6..18;21..34;53..62;79..86;89..95;108..113;128..149
425	2..11;29..46;47..55
426	5..26;33..68
427	1..11;19..38;38..49;52..60;90..98;100..105;129..137;155..160
428	23..31;47..61;65..72;87..94
429	3..12;31..39;59..69;82..87
430	81..91
431	36..44;83..89
432	28..42;56..76;110..117
433	5..14;43..49
434	9..17;15..21;61..70;80..89
435	1..24;32..40;52..60
436	1..11;17..30;29..43
437	20..30;28..35;40..50
438	10..28;76..85;91..99;107..112
439	86..92;104..110
440	36..42;47..57;63..71;106..121
441	19..25;27..44
442	1..14;21..29;36..42;44..51;62..81;101..109;111..119;138..149
443	10..18;25..31;33..40;51..70;89..94
444	3..8;19..26;32..44
445	1..11;19..38;38..49;52..60;130..139
446	12..20;28..37;43..66;90..102
447	15..20;24..31;36..47;68..82;88..96
448	29..45;83..91;88..94;132..144
449	22..33;54..64;86..96;102..108
450	27..39;47..60;101..107;155..164;270..281;287..300;306..312;327..332
451	13..24;60..70;77..83
452	8..14;74..82
453	6..12;63..78;77..91;97..103;102..108

454	32..44;66..72;101..114;166..174;209..235;243..252;258..263
455	69..76;131..139;164..173
456	54..63;95..103;187..202;211..216;249..261
457	14..21;31..45;80..88;187..194;347..353
458	47..62;79..86
459	1..8;27..37;90..97;99..106;123..140;145..163
460	8..17;35..45;131..139;162..169;175..180
461	1..6;13..23;58..66;89..101
462	44..53;86..93
463	62..70
464	50..57;59..69;67..73;79..95
465	10..17;23..44
466	3..15;71..78;110..121;125..131;259..269;296..306;312..318;340..346;353..363;370..379;407..412;417..425;448..453;483..493
467	5..12;20..30;70..78;82..100;106..115;129..135
468	8..16;22..31;36..45;75..84
469	14..23;98..105;106..116
470	11..23;26..31;54..62;101..107
471	23..29;66..81
472	23..29;93..100
473	8..25;79..89;103..109
474	37..45;80..89;94..101;125..130
475	37..45;80..89;94..101;125..130
476	7..26;23..36;36..45;78..83;80..85
477	45..53
478	1..7;16..22;78..93;96..102
479	24..33;41..50;61..80;93..100;129..136;160..170;199..208;267..276;325..335
480	5..14;43..51;102..116
481	2..15;16..24;53..62;87..97;100..109;109..133;145..152
482	2..15;16..24;53..62;87..97;100..109;109..133;145..152;168..176

Table VIII

Seq Id No	Chromosomal location
2	16p11-p13
12	16p11-p13
22	7q35-q36
25	chr.19
34	chr.17
35	6p21.3
40	chr.20
42	12p13.3
45	12p13.3
51	12p
56	22q11.2-q13.2
57	12p13
60	chr.10
62	chr.17
65	Xq13
67	chr.14
70	chr.7(1);7q11.23-q21.1(1)
71	chr.7(1);7q11.23-q21.1(1)
73	6p21.3
74	6p21.3
87	19q13.1
88	7q21-q22
94	17q11.2
99	6q21
101	6p11.2-p21.3
103	chr.17
106	6q15-q16.3
107	16p13.3
108	12q
113	1p33-p34.3
125	6p22.1-p22.3
126	16p13.3
127	14q11.2
135	22q11.2-q13.2
138	chr.3
141	12q24.1
146	3p21.3
147	chr.2
149	chr.17
150	21q
152	21q
154	20q12-q13.11
155	11p15.5
160	19q13.2
161	19q13.2
162	20q12-q13.11
164	21q
166	21q

170	6p12.1-p21.1
172	21q
173	chr.19
176	21q
177	21q
179	chr.6
183	chr.7
185	Xq21.3-q22.3
186	chr.20
192	11q12.2
195	chr.20
196	20q13.1-q13.2
197	7p15-p21
198	19q13.3
199	chr.2
201	Xq22.1-q23
202	Xq22.1-q23
204	chr.20
205	chr.5
206	chr.2
208	chr.5
214	chr.12
220	Xq28
224	chr.7
227	chr.14
230	chr.7
231	chr.7
238	19p13.3

Table IX

Seq Id No	Tissue distribution
1	Br:28;FB:25;FK:9;Ov:17;Pl:12;Pr:4;SC:2;SG:4;Te:9
2	Br:2;CP:1;FB:5;FK:1;Pl:3;Pr:10;SG:1
3	Br:1;CP:1;FB:33;FK:13;Li:2;Ov:19;PG:12;Pl:27;Pr:15;SG:9;SI:12
4	AG:1;CP:1;LG:1;Pr:3;Te:1
5	Pa:4;Pr:2
6	Li:1;Pa:4;Pr:3
7	Br:9;Pr:1;Te:3
8	Br:4;FB:1;Pr:3;SG:8
9	Br:4;Ce:1;Co:1;DM:4;FB:33;FK:16;He:3;Ki:6;LC:2;LG:4;Li:2;Lu:2;Ly:1;Ov:36;Pa:16;Pl:2;Pr:4;SC:2;SI:1;SN:1;Sp:1;UC:3;Ut:1
10	Br:1;CP:1;Pr:4;SG:2
11	Pr:2;SG:4
12	Br:1;CP:1;FB:5;FK:1;Pl:3;Pr:9;SG:1
13	FL:4;Li:4
14	Li:4;Te:3
15	Te:1
16	Li:3;Te:6
17	Ce:1;FB:6;Li:1;Pl:5;Te:16
18	Li:7;Te:6
19	Li:27;Te:9
20	Li:1;Te:3
21	Te:3
22	Te:3
23	Li:1;Te:6
24	Li:2;Te:2
25	Te:8
26	Te:5
27	LC:1;Te:2
28	Li:1;Te:2
29	AG:2;BM:1;Br:16;CP:1;Co:2;DM:1;FB:45;FK:62;FL:1;HP:3;LC:1;Li:2;Mu:1;Ov:2;Pr:10;SI:5;SN:3;Te:9;UC:1
30	Li:2
31	Br:3;CP:1;FB:1;FK:6;Pr:1;Te:2
32	Br:1;CP:1;Ce:6;Ov:1;Te:2
33	FK:5;SC:1
34	Br:1;FB:2;FK:48;Pl:2;SN:1
35	Te:1
36	FB:5;Pr:1;SN:1
37	FB:3;FK:1;Li:1;SG:5
38	FB:10
39	FB:3
40	Br:1;DM:1;FL:1;Pl:4;SG:13
41	FB:3;FK:1;Li:1;SG:5

42	BM:1;SG:19
43	SG:1
44	CP:1;FB:1;Mu:2;Pl:9;SG:7
45	BM:1;SG:20
46	BM:1;DM:1;FB:5;FK:6;FL:1;He:1;Ki:2;Ov:9;Pl:1;SG:1;SI:1;Te:1
47	Br:4;FB:4;Pr:3;SG:8
48	Br:12;Ce:1;Co:1;FB:5;FK:4;FL:5;HP:1;Ki:1;LC:1;Li:6;Ov:8;Pl:105;SC:1;SG:8;Te:4
49	Br:7;Ce:1;Co:1;FK:4;HP:1;Ki:1;LC:1;Li:5;Ov:8;Pl:5;SC:1;Te:1
50	AG:1;CP:4;Ce:1;DM:2;FB:6;FK:4;FL:2;HP:2;LC:1;LG:3;Li:31;Lu:3;Mu:1;Ov:25;Pl:15;Pr:20;SC:1;Te:75;UC:5;Ut:1
51	FL:1
52	Br:2;CP:1;FB:3;FK:1;FL:5;LC:2;Pl:1;Pr:2;UC:2
53	Br:3;FK:4;FL:4;HP:1;Li:3;Pl:11;SG:1;Te:1
54	Br:15;Ce:1;FB:10;FK:10;FL:1;He:1;Ki:6;LC:1;Li:4;Ov:32;Pa:3;Pl:2;Pr:4;SC:1;SN:2;Sp:4;Te:8;UC:1;Ut:1
55	FL:2
56	Br:1;FB:1;FL:1;Te:1
57	FL:4
58	FL:1;Li:1
59	FB:3;FK:1;Li:1;SG:5
60	Br:1;FB:1;FL:1;Pr:2
61	Br:2;Pl:1
62	Br:6;CP:1;Ce:7;FB:37;FK:4;FL:1;Pl:6;Pr:1;SG:3;SN:3;Te:1;UC:1
63	Br:10
64	Br:2;CP:2
65	Br:1;FB:11;LG:1;Th:1
66	Br:30;Ce:1;Co:1;FB:60;FK:15;FL:3;HP:1;Ki:1;LC:1;Li:6;Ov:57;PG:9;Pl:145;Pr:21;SC:1;SI:4;Te:4
67	Br:4;CP:1;FB:14;Ki:1;Li:1;Lu:2;Pr:1;Te:1
68	Br:10
69	AG:1;Br:48;FB:3;FK:5;HP:1;He:1;Li:1;Pl:11;SC:2;SG:1;Te:2;Ut:1
70	Br:11;DM:1;He:1
71	DM:1;He:1
72	Br:9;Pr:1;Te:2
73	Br:8;Pr:1
74	Br:5
76	AG:1;Br:49;FB:4;FK:5;HP:1;He:1;LC:1;Li:1;Pl:11;SC:2;SG:1;Te:2;Ut:1
77	Br:2;FK:2;HP:1;LC:1;Li:2;Ov:14;Pl:1;Pr:14;Te:5
78	Br:9;Ce:1;DM:1;FB:21;FK:18;FL:1;HP:1;He:1;Ki:9;LC:2;LG:4;Li:2;Lu:2;Ov:34;Pl:3;Pr:4;SC:1;SI:2;SN:2;Sp:1;Ut:1
79	Pr:1
80	Br:9;CP:2;Co:1;DM:6;FB:1;FK:6;He:2;Ki:4;LC:1;LG:1;Li:7;Ov:40;Pa:1;Pl:2;Pr:1;SN:2;Sp:1;Te:12;UC:1;Ut:3
81	FK:1;Te:1
82	Li:1

83	Br:2;CP:1;FB:10;FK:2;Ki:3;Li:7;Ov:10;SC:1;SN:1;Te:1;UC:1
84	Br:5;FB:14;FK:9;Li:6;Ov:17;SG:8;Te:8
85	Li:6;Te:2
86	Li:2;Te:2
87	Br:1;FB:35;FK:31;Li:20;Ov:37;PG:5;Pl:69;SI:5;Te:5
88	Li:1;Pr:1;Te:7;Ut:2
89	Te:1
90	Te:2
91	FB:15;FK:3;Li:2;Ov:17;Pr:4;SG:7;Te:4
92	Te:2
93	Br:4;FB:1;SN:1;Te:2
94	Te:1
95	Li:2
96	AG:1;Br:1;FB:1
97	FK:5;Te:2
98	Te:3
99	Br:3;FB:29;FK:1;Li:10;Ov:1;Pl:16;Pr:2;SG:1;Te:49
100	Br:2;FB:3;FK:1;Ov:3;Te:1
101	Br:10;FB:34;FK:1;Ov:1;Pl:85;Pr:1;Ut:1
102	FB:6
103	FB:6;Li:3;Pl:1;Pr:1;SG:1;Te:7
104	Br:26;CP:1;FB:8;FK:131;Pl:20;Pr:20
105	Br:3;CP:2;DM:2;FB:11;FK:3;LG:2;Ov:1;Pl:6;SC:2;SG:1;SN:4
106	FB:4
107	Br:3;FB:50;FK:59;FL:3;Pr:1
108	Br:1;FB:8;Li:1;Lu:1
109	FB:11;Pr:21
110	Br:14;Ce:1;FB:5;FK:5;FL:1;HP:1;He:2;Ki:3;LC:1;Li:4;Lu:1;Ov:7;Pl:2;Pr:1;SI:1;Sp:1;UC:1
111	Br:1;Ce:1;FK:1;HP:1;He:2;Ki:3;LC:1;Li:3;Lu:1;Ov:7;Sp:1;UC:1
112	Br:1;HP:1;Lu:1;Pr:1;SG:9;Ut:1
113	HP:1;SG:4
114	FK:9
115	AG:1;Br:3;CP:1;FB:14;FK:19;FL:1;HP:1;Pr:1;SG:1
116	Br:5;CP:1;Ce:1;Co:1;DM:5;FK:3;FL:1;LC:3;LG:1;Lu:1;Ov:23;Pl:1;Te:8;UC:2;Ut:4
117	Br:1;Ce:1;FB:1;FK:1;FL:2;Pl:3;SN:1;Te:1;UC:1
118	CP:1;DM:1;FB:5;FK:2;FL:2;He:1;Lu:1;Ly:1;Ov:23;Pr:1;SN:2;Sp:2;Ut:1
119	Li:2;Te:7
120	Br:6;Co:2;FB:1;FK:6;FL:1;Ov:3;Pl:32;Pr:1;SN:1
121	AG:1;Br:4
122	Br:5
123	Br:1
124	Br:2;Ki:2;Li:1;Ov:7;UC:1
125	Br:2;FB:1;FL:6;He:1;Li:1;Ov:1;Pl:2;Pr:10;Te:1;Th:1
126	Br:1

127	BM:1;Br:2;CP:2;FB:1;FK:3;HP:1;He:1;LG:1;Pl:1;Pr:1;SC:2;SG:2;Te:5;Ut:3
128	Br:1
129	Br:2;FB:6;Li:1;SG:3;Te:2
130	Br:25;FB:3;FL:2
131	Br:1
132	Br:1
133	Br:1
134	Br:2;SN:1
135	Br:1
136	AG:1;Br:1;FL:1
137	Br:1;Ce:1;FB:1;FK:1;FL:2;Pl:3;SN:1;Te:1;UC:1
138	Br:43
139	Br:11;CP:2;Co:1;DM:6;FB:1;FK:6;He:2;Ki:4;LC:1;LG:1;Ov:40;Pa:1;Pl:2;Pr:1;SN:2;Sp:1;Te:9;UC:1;Ut:3
140	Br:23;Ce:1;DM:3;FB:38;FK:17;FL:2;HP:1;He:1;Ki:8;LC:3;LG:2;Li:6;Lu:1;Ly:1;Ov:40;Pr:4;SC:2;SN:4;Sp:1;Te:5;UC:1;Ut:1
141	Br:39;FB:3;SN:2
142	Br:10;SN:2
143	Br:26;FK:2;HP:1;LC:1;Li:2;Ov:14;Pl:3;Pr:3;Te:5
144	Br:14;Pr:2
145	FB:12;LG:1;Pr:4;Te:1;Ut:2
146	Li:1;Ov:2;Pr:5;SG:11
147	Li:1;Te:1
148	Br:1;FB:1;Li:1;Te:1
149	Br:3;FB:5;FK:5;Li:1;Pl:8;Te:5
150	FK:6;Pr:2;SG:8
151	FK:9
152	FK:6;Pr:2;SG:9
153	Te:1
154	FB:28;Ov:4
155	Br:21;Ce:1;FB:32;FK:4
156	Br:5;CP:1;FB:16;FK:3;He:1;Ki:5;Li:1;Ov:15;Pl:3;SG:2;SI:1;Sp:1;UC:1
157	FB:14;FK:1;FL:1;SG:1
158	FB:7
159	FB:10
160	Ce:2;FB:12
161	Ce:2
162	FB:28;Ov:2
163	FB:14;FK:1;FL:1;SG:1
164	FK:4;Pr:1;SG:9
165	Br:4;Co:1;Ki:1;Ov:2;Pr:4;SG:1
166	FK:6;Pr:2;SG:9
167	Br:1;FB:1;SG:5
168	Br:1;FB:5;FK:7;SG:1;UC:1
169	FK:2
170	FL:12

171	Br:2;FB:1;FK:1;Pl:7
172	Br:106;FB:2;Pl:7
173	Br:14;FB:1;Pl:2;Te:1
174	Br:17;He:1;Pl:1;SC:2;Te:1
175	Br:14;Pr:2
176	Br:106;FB:2;Pl:7
177	Br:114;FB:7;FK:7;Ov:2;Pl:7;Pr:2;Te:9
178	Br:16;CP:2;FB:2;FK:2;FL:1;Li:1;Pl:13;Pr:3;SC:1;Ut:1
179	FL:1;HP:2;Pr:2;Te:1
180	Pr:2
181	FB:1;Ov:2;Pr:1;UC:1
182	BM:1;Br:4;DM:1;FB:6;FK:6;Ki:5;LC:2;LG:1;Li:1;Lu:1;Ov:15;Pl:1;Pr:2;SC:1;Sp:2;Te:2;Ut:1
183	Br:8;CP:1;Co:2;DM:4;FB:1;FK:1;Ki:4;LC:1;Li:3;Ov:33;Pl:1;Pr:5;SC:2;SN:1;Sp:1;Te:5;UC:1;Ut:2
184	Pr:1
185	FB:2;Li:1;Ov:1;SG:7;Te:5
186	Te:3
187	Te:1
188	Br:18;CP:1;DM:5;FB:40;FK:23;FL:2;He:3;Ki:10;LC:2;LG:1;Li:13;Lu:3;Ly:2;Mu:1;Ov:54;Pl:5;Pr:14;SC:2;SG:2;SI:2;SN:4;Sp:3;Te:4;UC:4
189	Li:1;Te:1
190	Br:7;CP:1;FB:1;FK:4;FL:5;He:1;Li:1;Ov:1;Pl:2;Pr:4;SG:1
191	Li:2;Te:4
192	AG:1;Br:2;CP:1;FB:32;FK:1;Li:1;Ov:36;Pl:49;Pr:3;SC:1;SG:4;SN:4;Te:9;UC:1;Ut:2
193	FB:31;FK:75;FL:7;Ov:12;Pl:23;Pr:8;SG:3;Te:16
194	Te:2
195	Te:7
196	Te:2
197	Te:3
198	Li:10;Te:43
199	Br:35;CP:3;FB:39;FK:56;FL:7;HP:1;LG:1;Li:1;Ly:1;Ov:2;Pl:10;Pr:8;SG:1;Te:4;Ut:2
200	FB:17;FK:9;FL:5;Ov:21;Pl:41;Te:3
201	FK:16;SI:1
202	Br:1;Co:1;FB:111;FK:25;He:1;Li:4;Ov:3;Pr:6;Te:1
204	Te:7
205	Li:7;Te:28
206	FB:28;Li:2;Ov:23;PG:11;Pl:45;SG:17;SI:11;Te:9
207	FB:16;FK:1;Ov:1;SC:1;Te:1
208	FB:5
209	FB:6
210	Br:1;FB:22
211	Br:2;Ce:3;FB:6;FK:1
212	Br:1;Co:2;FB:22;FK:2;LG:2;Mu:2;Pl:2;SG:4

213	Br:2;DM:1;FB:8;FK:8;FL:1;Ki:1;LG:3;Ov:5;Pa:1;Pl:4;Pr:1;SN:2;UC:1
214	FB:7
215	FB:4
216	Ov:3;SG:3
217	Br:4;CP:2;DM:1;FB:9;FK:3;Ki:2;LC:1;LG:1;Lu:3;Ly:1;Ov:14;Pl:1;Pr:1;SC:1;SG:2;Sp:1;Te:1;Ut:1
218	FB:4;FK:2;Pl:1;Pr:11;SG:1
219	Br:7;CP:3;FB:2;FL:1;HP:4;Lu:1;Ly:2;Mu:1;Ov:3;Pl:1;Pr:1;SN:2;Te:1
220	Br:1;FL:1;Pl:2
221	Co:1;FB:2;FL:1;Li:1;Pl:2
222	FL:1;SG:2
223	Li:1;Te:1
225	Li:10
226	Li:1;Te:4
227	Li:1
228	Br:1
229	Br:3
230	Br:5;Ce:1;Co:1;DM:3;FB:1;FK:1;He:1;LC:1;LG:2;Ov:16;Pl:3;Pr:1;Te:2;Ut:1
231	Br:3;Ce:1;Co:1;DM:3;FB:1;FK:1;He:1;LC:1;LG:2;Ov:16;Pl:3;Pr:1;Te:2;Ut:1
232	AG:1;Br:17;CP:2;DM:1;FB:51;FK:9;FL:3;Li:3;Ov:3;Pl:2;Pr:10;SC:1;SG:5;Te:2;Ut:1
233	Br:13
234	Br:5
235	Br:1;Pl:1
236	Br:9
237	Br:22;DM:2;FB:17;FK:9;Ki:4;LG:1;Li:1;Lu:2;Ov:24;Pr:3;SC:1;SI:1;SN:2;Te:2
238	Br:17
239	Br:11
240	Br:28;Ce:1;DM:5;FB:52;FK:40;FL:2;HP:1;He:2;Ki:3;LC:1;LG:3;Li:1;Ly:1;Ov:28;Pl:1;Pr:5;SC:1;SI:1;SN:3;Sp:6;Te:1;UC:1;Ut:1
241	Br:4;Ce:1;DM:5;FB:5;FK:7;HP:1;He:2;Ki:3;LC:1;LG:3;Li:1;Ly:1;Ov:28;Pl:1;SC:1;SN:3;Sp:6;Te:1;UC:1;Ut:1

Table X

Seq Id No	Low frequency expression	High frequency expression
1	-	Br,Ov
2	-	Pr
3	Br,Te	Ov,PG,Pl,SI
4	-	AG
5	-	Pa
6	-	Pa
7	-	Br
8	-	SG
9	Br,Te	DM,He,Ki,Ov,Pa
10	-	Pr
11	-	SG
12	-	Pr
13	-	FL,Li
14	-	Li,Te
15	-	Te
16	-	Li,Te
17	-	Te
18	-	Li,Te
19	-	Li,Te
20	-	Te
21	-	Te
22	-	Te
23	-	Te
24	-	Li
25	-	Te
26	-	Te
27	-	LC,Te
28	-	Te
29	Pl	FK
30	-	Li
31	-	FK
32	-	Ce
33	-	FK,SC
34	FB	FK
35	-	Te
36	-	SN
37	-	SG
38	-	FB
40	-	SG
41	-	SG
42	-	BM,SG
43	-	SG

44	-	Mu,Pl,SG
45	-	BM,SG
46	-	BM,Ki,Ov
47	-	SG
48	FB,FK,Pr	Pl
49	-	Ki,Ov
50	Br,FB,FK,SG	Li,Ov,Te
51	-	FL
52	-	FL,LC,UC
53	-	Pl
54	-	Ki,Ov,Pa,Sp
55	-	FL
57	-	FL
58	-	FL
59	-	SG
62	-	Ce,FB
63	-	Br
64	-	CP
65	-	FB,Th
66	FK,SG,Te	Ov,PG,Pl
67	-	FB,Ki,Lu
68	-	Br
69	FB	Br
70	-	Br,DM,He
71	-	DM,He
72	-	Br
73	-	Br
74	-	Br
75	-	Br
76	FB	Br
77	FB	Ov,Pr
78	-	Ki,Ov
80	FB	DM,Ki,Ov
82	-	Li
83	-	Ki,Li,Ov
84	-	Ov
85	-	Li
86	-	Li
87	Br,Pr,SG	Ov,PG,Pl
88	-	Te,Ut
89	-	Te
90	-	Te
91	-	Ov
92	-	Te
93	-	SN
94	-	Te

95	-	Li
96	-	AG
97	-	FK
98	-	Te
99	FK	Te
100	-	Ov
101	FK	Pl
102	-	FB
103	-	Te
104	FB,Li,SG,Te	FK
105	-	DM,SN
106	-	FB
107	Br,Pl	FB,FK
108	-	FB,Lu
109	-	Pr
110	-	He,Ki,Ov
111	-	Ce,He,Ki,Lu,Ov
112	-	Lu,SG
113	-	HP,SG
114	-	FK
115	-	FK
116	FB	DM,LC,Ov,Ut
117	-	Ce,UC
118	-	Ov,Sp
119	-	Te
120	FB	Co,Pl
121	-	AG,Br
122	-	Br
124	-	Ki,Ov
125	-	FL,Pr,Th
127	-	BM,SC,Ut
130	-	Br
134	-	SN
136	-	AG
137	-	Ce,UC
138	FB	Br
139	FB	DM,Ki,Ov,Ut
140	Pl	Ki,Ov
141	-	Br
142	-	Br,SN
143	FB	Br,Ov
144	-	Br
145	-	FB,Ut
146	-	SG
149	-	Pl
150	-	FK,SG

151	-	FK
152	-	FK,SG
153	-	Te
154	-	FB,Ov
155	-	Br,FB
156	-	Ki,Ov
157	-	FB
158	-	FB
159	-	FB
160	-	Ce,FB
161	-	Ce
162	-	FB
163	-	FB
164	-	SG
165	-	Co,Ki,Ov
166	-	FK,SG
167	-	SG
168	-	FK
169	-	FK
170	-	FL
171	-	Pl
172	FB,FK,Pr	Br
173	-	Br
174	-	Br,He,SC
175	-	Br
176	FB,FK,Pr	Br
177	FB	Br
178	-	Br,Pl
179	-	HP
180	-	Pr
181	-	Ov,UC
182	-	Ki,Ov,Sp
183	FB	DM,Ki,Ov
185	-	SG,Te
186	-	Te
187	-	Te
188	Pl	DM,Ki,Ov
190	-	FL
191	-	Te
192	Br,FK	Ov,Pl
193	Br	FK,Ov
194	-	Te
195	-	Te
196	-	Te
197	-	Te
198	FB	Li,Te

199	-	FK
200	Br	Ov,Pl
201	-	FK
202	-	FK
203	Br,Pl	FB
204	-	Te
205	-	Li,Te
206	Br,FK,Pr	Ov,PG,Pl,SG,SI
207	-	FB
208	-	FB
209	-	FB
210	-	FB
211	-	Ce
212	-	Co,FB,Mu
213	-	Ki,LG,Ov
214	-	FB
215	-	FB
216	-	Ov,SG
217	-	Ki,Lu,Ov
218	-	Pr
219	-	CP,HP,Ly,Ov,SN
221	-	Co
222	-	SG
223	-	SG
225	-	Li
226	-	Te
227	-	Li
229	-	Br
230	-	DM,Ov
231	-	DM,Ov
232	-	FB
233	-	Br
234	-	Br
236	-	Br
237	-	Ki,Lu,Ov
238	-	Br
239	-	Br
240	Pl,Te	DM,FK,Ki,Ov,Sp
241	FB	DM,He,Ki,Ov,Sp

Table XI

Seq Id No	Subcellular localization
7	nuclear
13	extracellular, including cell wall
20	mitochondrial
21	nuclear
26	nuclear
35	nuclear
37	endoplasmic reticulum
38	extracellular, including cell wall
39	endoplasmic reticulum
41	endoplasmic reticulum
59	endoplasmic reticulum
70	nuclear
71	nuclear
72	nuclear
78	nuclear
98	nuclear
99	nuclear
105	mitochondrial
108	endoplasmic reticulum
116	mitochondrial
117	mitochondrial
134	nuclear
135	nuclear
137	mitochondrial
159	nuclear
160	nuclear
161	nuclear
171	nuclear
178	endoplasmic reticulum
182	nuclear
184	nuclear
185	endoplasmic reticulum
186	nuclear
187	nuclear
188	nuclear
194	nuclear
195	nuclear
196	nuclear
200	mitochondrial
204	nuclear
205	nuclear
206	nuclear

211	nuclear
212	nuclear
213	nuclear
214	endoplasmic reticulum
215	endoplasmic reticulum
216	endoplasmic reticulum
218	nuclear
220	endoplasmic reticulum
224	nuclear
225	nuclear
230	mitochondrial
231	mitochondrial
238	cytoplasmic

Table XII

Seq Id No in priority applications	Internal designation	Seq Id No in present application
119	119-003-4-0-C2-CS	1
220	105-016-1-0-D3-CS	2
345	105-016-3-0-G10-CS	3
334	105-026-1-0-A5-CS	4
159	105-031-1-0-A11-CS	5
219	105-031-2-0-D3-CS	6
250	105-035-2-0-C6-CS	7
217	105-037-2-0-H11-CS	8
340	105-053-4-0-E8-CS	9
115	105-074-3-0-H10-CS	10
31	105-089-3-0-G10-CS	11
198	105-095-2-0-G11-CS	12
154	106-006-1-0-E3-CS	13
366	106-037-1-0-E9-CS.cor	14
366	106-037-1-0-E9-CS.fr	15
79	106-043-4-0-H3-CS	16
95	110-007-1-0-C7-CS	17
364	114-016-1-0-H8-CS	18
246	116-004-3-0-A6-CS	19
187	116-054-3-0-E6-CS	20
203	116-055-1-0-A3-CS	21
298	116-055-2-0-F7-CS	22
277	116-088-4-0-A9-CS	23
41	116-091-1-0-D9-CS	24
353	116-110-2-0-F4-CS	25
78	116-111-1-0-H9-CS	26
245	116-111-4-0-B3-CS	27
104	116-115-2-0-F8-CS	28
259	116-119-3-0-H5-CS	29
269	117-001-5-0-G3-CS	30
166	145-25-3-0-B4-CS.cor	31
166	145-25-3-0-B4-CS.fr	32
169	145-56-3-0-D5-CS	33
312	145-59-2-0-A7-CS	34
273	157-15-4-0-B11-CS	35
190	160-103-1-0-F11-CS	36
244	160-37-2-0-H7-CS	37
151	160-58-3-0-H3-CS	38
149	160-75-4-0-A9-CS	39
307	174-10-2-0-F8-CS	40
264	174-33-3-0-F6-CS	41

168	174-38-1-0-B6-CS	42
202	174-38-3-0-C9-CS	43
28	174-39-2-0-A3-CS	44
331	174-41-1-0-A6-CS	45
258	174-5-3-0-H7-CS	46
84	174-7-4-0-H1-CS	47
294	175-1-3-0-E5-CS.cor	48
294	175-1-3-0-E5-CS.fr	49
310	180-19-4-0-F4-CS	50
311	181-10-1-0-D10-CS	51
263	181-16-1-0-G7-CS	52
304	181-16-2-0-A7-CS	53
109	181-20-3-0-B5-CS	54
121	181-3-3-0-B8-CS	55
181	181-3-3-0-C9-CS	56
191	182-1-2-0-D12-CS	57
193	184-1-4-0-C11-CS	58
192	184-4-1-0-A11-CS	59
116	187-12-4-0-A8-CS	60
268	187-2-2-0-A3-CS	61
123	187-31-0-0-f12-CS	62
234	187-34-0-0-l12-CS	63
185	187-37-0-0-c10-CS	64
279	187-38-0-0-l10-CS	65
114	187-39-0-0-k12-CS	66
211	187-41-0-0-i21-CS	67
236	188-11-1-0-B3-CS	68
35	188-18-4-0-A9-CS	69
299	188-28-4-0-B12-CS.cor	70
299	188-28-4-0-B12-CS.fr	71
72	188-28-4-0-D4-CS	72
242	188-41-1-0-B8-CS.cor	73
242	188-41-1-0-B8-CS.fr	74
173	188-45-1-0-D9-CS	75
106	188-9-2-0-E1-CS	76
130	105-079-3-0-A11-CS	77
323	105-092-1-0-H7-CS	78
160	105-141-4-0-H9-CS	79
272	109-013-1-0-B9-CS	80
226	110-008-4-0-D9-CS	81
333	114-001-3-0-A2-CS	82
315	114-028-2-0-C1-CS	83
300	114-032-1-0-H10-CS	84
57	114-043-2-0-A10-CS	85
137	114-044-1-0-C5-CS	86
107	116-003-3-0-D10-CS	87

164	116-003-3-0-G12-CS	88
108	116-011-2-0-F11-CS	89
101	116-033-3-0-E4-CS	90
157	116-041-4-0-B6-CS	91
75	116-044-2-0-C4-CS	92
322	116-075-1-0-E6-CS	93
124	116-094-4-0-G5-CS	94
289	117-005-3-0-F2-CS	95
122	121-007-3-0-D9-CS	96
208	145-91-3-0-D10-CS	97
282	157-17-1-0-F4-CS	98
129	160-11-3-0-G8-CS	99
317	160-24-1-0-F12-CS	100
308	160-24-2-0-E9-CS	101
25	160-25-4-0-D2-CS	102
243	160-31-3-0-A11-CS	103
346	160-32-1-0-F6-CS	104
60	160-37-1-0-A3-CS	105
305	160-40-3-0-E9-CS	106
48	160-58-3-0-E4-CS	107
238	160-85-3-0-D4-CS	108
251	160-95-3-0-A11-CS	109
196	162-10-4-0-F9-CS.cor	110
196	162-10-4-0-F9-CS.fr	111
347	174-13-2-0-E4-CS	112
77	174-46-2-0-B11-CS	113
188	179-8-2-0-A6-CS	114
235	180-22-3-0-B6-CS	115
45	181-13-1-0-F7-CS	116
265	181-15-4-0-F7-CS	117
280	181-20-1-0-G7-CS	118
281	184-15-3-0-D1-CS	119
39	187-12-2-0-G11-CS	120
165	187-2-2-0-A12-CS	121
326	187-30-0-0-k23-CS	122
330	187-36-0-0-e19-CS	123
368	187-38-0-0-d22-CS	124
71	187-39-0-0-b9-CS	125
224	187-39-0-0-g6-CS	126
90	187-45-0-0-l18-CS	127
216	187-45-0-0-m21-CS	128
83	187-45-0-0-n8-CS	129
342	187-46-0-0-f23-CS	130
262	187-5-1-0-A12-CS	131
257	187-5-1-0-F6-CS	132
293	187-5-2-0-B2-CS	133

231	187-5-3-0-D5-CS	134
287	187-51-0-0-f9-CS	135
325	187-6-1-0-B9-CS	136
309	187-6-4-0-C10-CS	137
359	188-19-2-0-C8-CS	138
68	188-22-4-0-G6-CS	139
233	188-28-4-0-D11-CS	140
369	188-29-1-0-E10-CS	141
155	188-34-4-0-E5-CS	142
327	188-9-3-0-A5-CS	143
283	105-021-3-0-C3-CS	144
29	105-037-4-0-H12-CS	145
100	105-073-2-0-A7-CS	146
99	109-002-4-0-C6-CS	147
360	109-003-1-0-G4-CS	148
321	116-118-4-0-A8-CS	149
120	145-52-2-0-D12-CS	150
230	145-7-2-0-G5-CS	151
177	145-7-3-0-D3-CS	152
43	157-17-2-0-C1-CS	153
352	160-101-3-0-H2-CS	154
47	160-12-1-0-D10-CS	155
195	160-28-4-0-C4-CS	156
344	160-31-3-0-E4-CS	157
61	160-40-1-0-H4-CS	158
237	160-54-1-0-F7-CS	159
32	160-88-3-0-A8-CS.cor	160
32	160-88-3-0-A8-CS.fr	161
97	160-99-4-0-E4-CS	162
249	161-5-4-0-B6-CS	163
218	174-17-1-0-D6-CS	164
266	174-32-4-0-F8-CS	165
161	174-38-4-0-D11-CS	166
113	174-8-2-0-C10-CS	167
255	179-14-2-0-F11-CS	168
24	179-9-4-0-B8-CS	169
128	181-10-1-0-C9-CS	170
58	187-5-3-0-C7-CS	171
358	188-26-4-0-F5-CS	172
171	188-27-3-0-G1-CS	173
98	188-29-2-0-H1-CS	174
133	188-31-1-0-E6-CS	175
49	188-45-1-0-D3-CS	176
42	188-5-1-0-H6-CS	177
148	188-9-1-0-C10-CS	178
319	105-016-3-0-C5-CS	179

254	105-026-4-0-D9-CS	180
55	105-053-2-0-D9-CS	181
89	105-069-3-0-A11-CS	182
212	105-076-4-0-F6-CS	183
143	105-135-2-0-F9-CS	184
179	106-023-4-0-F6-CS	185
54	110-001-3-0-C11-CS	186
354	110-002-3-0-F9-CS	187
297	114-019-3-0-D9-CS	188
291	114-029-1-0-C6-CS	189
30	114-032-4-0-B1-CS	190
363	114-070-2-0-H4-CS	191
194	116-016-3-0-F11-CS	192
76	116-022-4-0-G2-CS	193
232	116-052-2-0-H8-CS	194
139	116-053-4-0-B4-CS	195
145	116-094-3-0-H2-CS	196
206	116-112-4-0-C7-CS	197
88	116-123-3-0-F12-CS	198
69	123-008-1-0-C5-CS	199
285	145-53-2-0-H8-CS	200
96	145-57-2-0-C9-CS.cor	201
96	145-57-2-0-C9-CS.fr	202
26	145-7-3-0-B12-CS	203
261	157-12-2-0-D1-CS	204
349	157-16-2-0-D5-CS	205
92	157-18-2-0-A7-CS	206
81	160-103-1-0-B10-CS	207
74	160-104-4-0-F3-CS	208
66	160-22-2-0-D10-CS	209
270	160-24-3-0-F12-CS	210
318	160-3-2-0-H3-CS	211
215	160-58-2-0-A2-CS	212
214	160-73-1-0-B4-CS	213
94	160-75-4-0-E6-CS	214
73	160-97-3-0-E9-CS	215
296	174-1-4-0-E9-CS	216
127	174-12-4-0-C2-CS	217
350	180-19-4-0-H2-CS	218
320	181-10-4-0-G12-CS	219
292	181-3-2-0-F6-CS	220
125	181-4-4-0-A12-CS	221
91	181-9-2-0-F12-CS.cor	222
91	181-9-2-0-F12-CS.fr	223
338	184-13-3-0-E11-CS	224
102	184-4-2-0-D3-CS	225

105	184-7-1-0-E7-CS	226
112	184-8-4-0-G9-CS	227
267	187-10-3-0-G9-CS	228
260	187-32-0-0-m20-CS	229
343	187-32-0-0-n21-CS.cor	230
343	187-32-0-0-n21-CS.fr	231
341	187-4-2-0-E6-CS	232
59	187-40-0-0-i15-CS	233
111	187-47-0-0-g24-CS	234
40	187-9-3-0-A2-CS	235
355	188-26-4-0-H1-CS	236
37	188-35-3-0-G9-CS	237
180	188-38-4-0-D8-CS	238
152	188-41-1-0-E6-CS	239
93	188-42-2-0-F3-CS.cor	240
93	188-42-2-0-F3-CS.fr	241

WHAT IS CLAIMED IS:

1. An isolated polynucleotide, said polynucleotide comprising a nucleic acid sequence encoding:

i) a polypeptide comprising an amino acid sequence having at least about 80% identity to any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool; or

ii) a biologically active fragment of said polypeptide.

2. The polynucleotide of claim 1, wherein said polypeptide comprises any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of the polypeptides encoded by the clone inserts of the deposited clone pool.

3. The polynucleotide of claim 1, wherein said polypeptide comprises a signal peptide.

4. The polynucleotide of claim 1, wherein said polypeptide is a mature protein.

5. The polynucleotide of claim 1, wherein said nucleic acid sequence has at least about 80% identity over at least about 100 contiguous nucleotides to any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool.

6. The polynucleotide of claim 1, wherein said polynucleotide hybridizes under stringent conditions to a polynucleotide comprising any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool.

7. The polynucleotide of claim 5, wherein said nucleic acid sequence comprises any one of the sequences shown as SEQ ID NOs:1-241 or any one the sequences of the clone inserts of the deposited clone pool.

8. The polynucleotide of claim 1, wherein said polynucleotide is operably linked to a promoter.

9. An expression vector comprising the polynucleotide of claim 8.

10. A host cell recombinant for the polynucleotide of claim 1.

11. A non-human transgenic animal comprising the host cell of claim 10.

12. A method of making a GENSET polypeptide, said method comprising

a) providing a population of host cells comprising the polynucleotide of claim 8; and

b) culturing said population of host cells under conditions conducive to the production of said polypeptide within said host cells.

13. The method of claim 12, further comprising purifying said polypeptide from said population of host cells.

14. A method of making a GENSET polypeptide, said method comprising

a) providing a population of cells comprising the polynucleotide of claim 8;

b) culturing said population of cells under conditions conducive to the production of said polypeptide within said cells; and

c) purifying said polypeptide from said population of cells.

15. An isolated polynucleotide, said polynucleotide comprising a nucleic acid sequence having at least about 80% identity over at least about 100 contiguous nucleotides to any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool.

16. The polynucleotide of claim 15, wherein said polynucleotide hybridizes under stringent conditions to a polynucleotide comprising any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool.

17. The polynucleotide of claim 15, wherein said polynucleotide comprises any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool.

18. A biologically active polypeptide encoded by the polynucleotide of claim 15.

19. An isolated polypeptide or biologically active fragment thereof, said polypeptide comprising an amino acid sequence having at least about 80% sequence identity to any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool.

20. The polypeptide of claim 19, wherein said polypeptide is selectively recognized by an antibody raised against an antigenic polypeptide, or an antigenic fragment thereof, said antigenic polypeptide comprising any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool.

21. The polypeptide of claim 19, wherein said polypeptide comprises any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool.

22. The polypeptide of claim 19, wherein said polypeptide comprises a signal peptide.

23. The polypeptide of claim 19, wherein said polypeptide is a mature protein.

24. An antibody that specifically binds to the polypeptide of claim 19.

25. A method of determining whether a GENSET gene is expressed within a mammal, said method comprising the steps of:

a) providing a biological sample from said mammal

b) contacting said biological sample with either of:

i) a polynucleotide that hybridizes under stringent conditions to the polynucleotide of claim 1; or

ii) a polypeptide that specifically binds to the polypeptide of claim 19; and

c) detecting the presence or absence of hybridization between said polynucleotide and an RNA species within said sample, or the presence or absence of binding of said polypeptide to a protein within said sample;

wherein a detection of said hybridization or of said binding indicates that said GENSET gene is expressed within said mammal.

26. The method of claim 25, wherein said polynucleotide is a primer, and wherein said hybridization is detected by detecting the presence of an amplification product comprising the sequence of said primer.

27. The method of claim 25, wherein said polypeptide is an antibody.

28. A method of determining whether a mammal has an elevated or reduced level of GENSET gene expression, said method comprising the steps of :

a) providing a biological sample from said mammal; and

b) comparing the amount of the polypeptide of claim 19, or of an RNA species encoding said polypeptide, within said biological sample with a level detected in or expected from a control sample;

wherein an increased amount of said polypeptide or said RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has an elevated level of said GENSET gene expression, and wherein a decreased amount of said polypeptide or said RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has a reduced level of said GENSET gene expression.

29. A method of identifying a candidate modulator of a GENSET polypeptide, said method comprising :

a) contacting the polypeptide of claim 18 with a test compound; and

b) determining whether said compound specifically binds to said polypeptide;

wherein a detection that said compound specifically binds to said polypeptide indicates that said compound is a candidate modulator of said GENSET polypeptide.

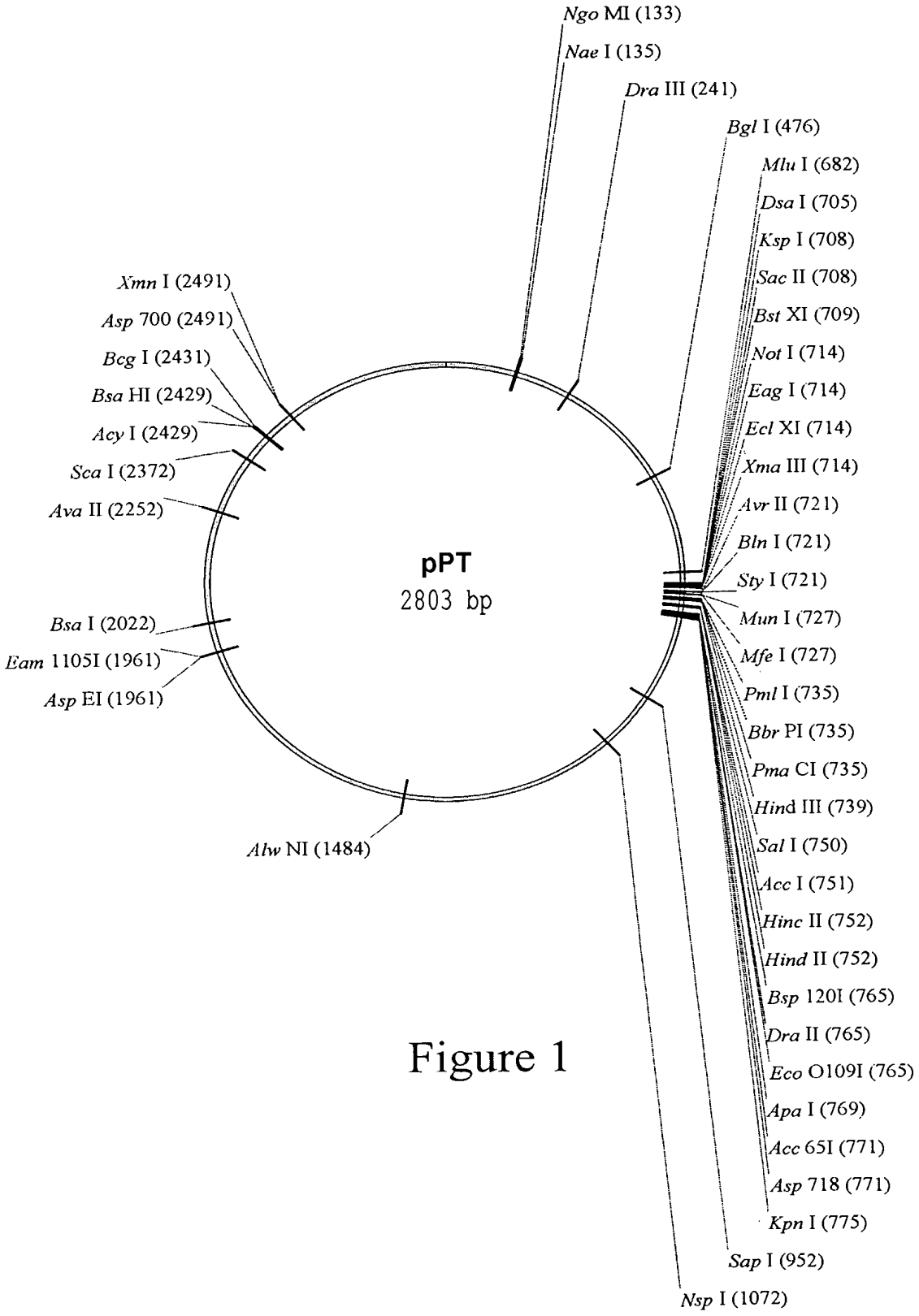
Ngo MI (133)
Nae I (135)
Dra III (241)
Bgl I (476)
Mlu I (682)
Dsa I (705)
Ksp I (708)
Sac II (708)
Bst XI (709)
Not I (714)
Eag I (714)
Ecl XI (714)
Xma III (714)
Avr II (721)
Bln I (721)
Sty I (721)
Mun I (727)
Mfe I (727)
Pml I (735)
Bbr PI (735)
Pma CI (735)
Hind III (739)
Sal I (750)
Acc I (751)
Hinc II (752)
Hind II (752)
Bsp 120I (765)
Dra II (765)
Eco O109I (765)
Apa I (769)
Acc 65I (771)
Asp 718 (771)
Kpn I (775)
Sap I (952)
Nsp I (1072)
Alw NI (1484)
Asp EI (1961)
Eam 1105I (1961)
Bsa I (2022)
Ava II (2252)
Sca I (2372)
Acy I (2429)
Bsa HI (2429)
Bcg I (2431)
Asp 700 (2491)
Xmn I (2491)
pPT
2803 bp

Figure 1

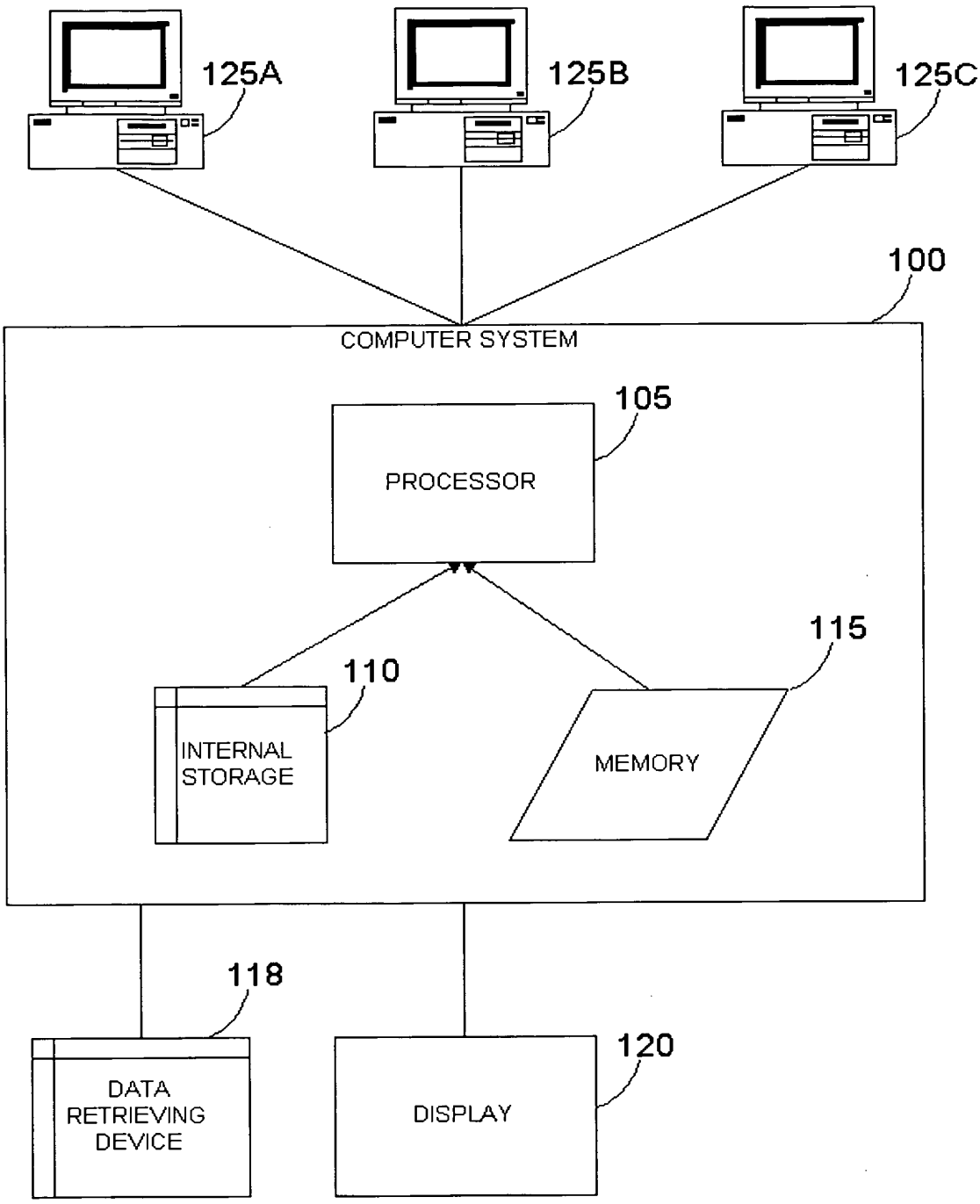
125A
125B
125C
100
COMPUTER SYSTEM
105
PROCESSOR
115
110
INTERNAL STORAGE
MEMORY
118
DATA RETRIEVING DEVICE
120
DISPLAY

Figure 2

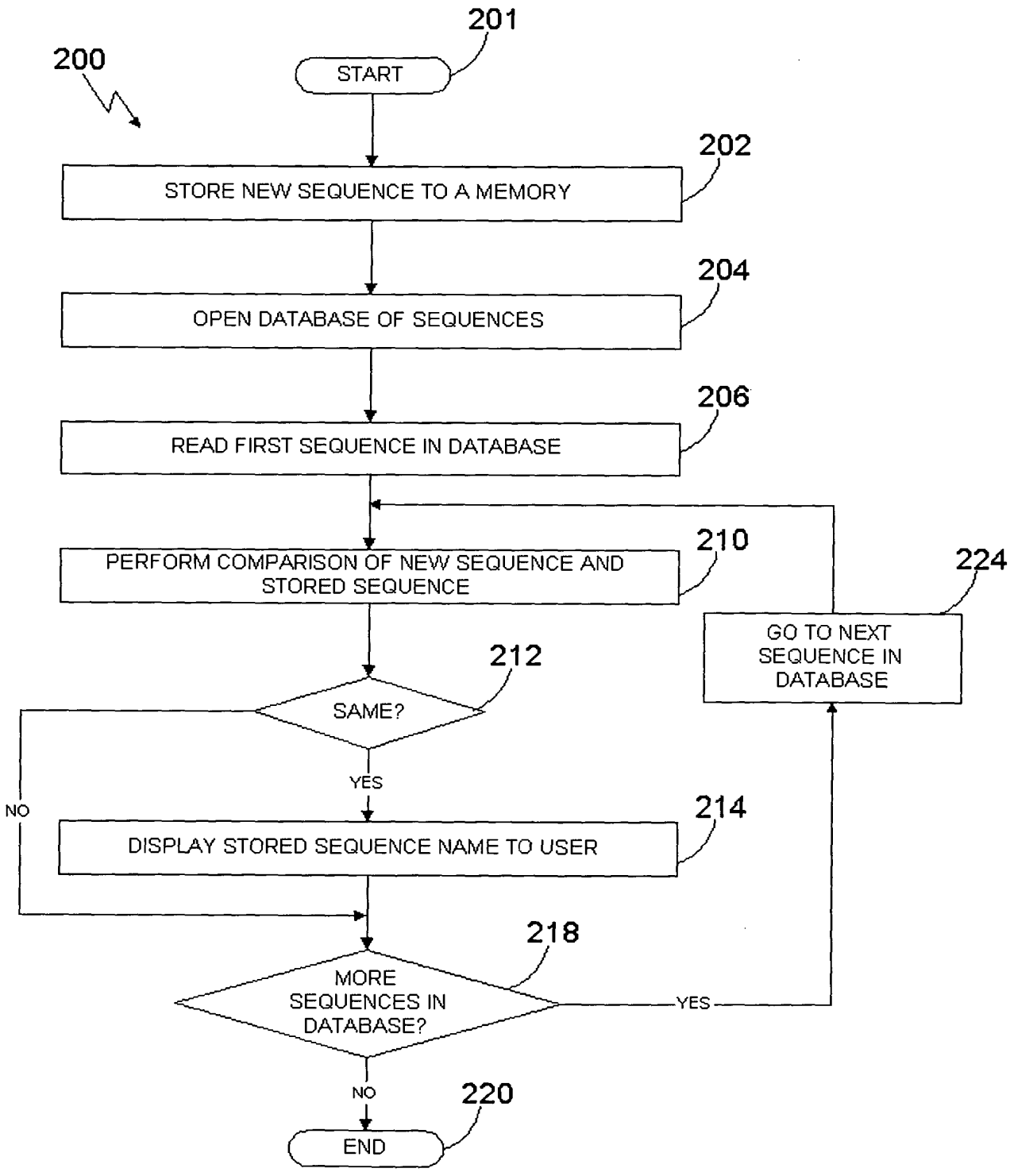
200
201
START
202
STORE NEW SEQUENCE TO A MEMORY
204
OPEN DATABASE OF SEQUENCES
206
READ FIRST SEQUENCE IN DATABASE
210
PERFORM COMPARISON OF NEW SEQUENCE AND STORED SEQUENCE
212
SAME?
YES
NO
214
DISPLAY STORED SEQUENCE NAME TO USER
218
MORE SEQUENCES IN DATABASE?
YES
NO
220
END
224
GO TO NEXT SEQUENCE IN DATABASE

Figure 3

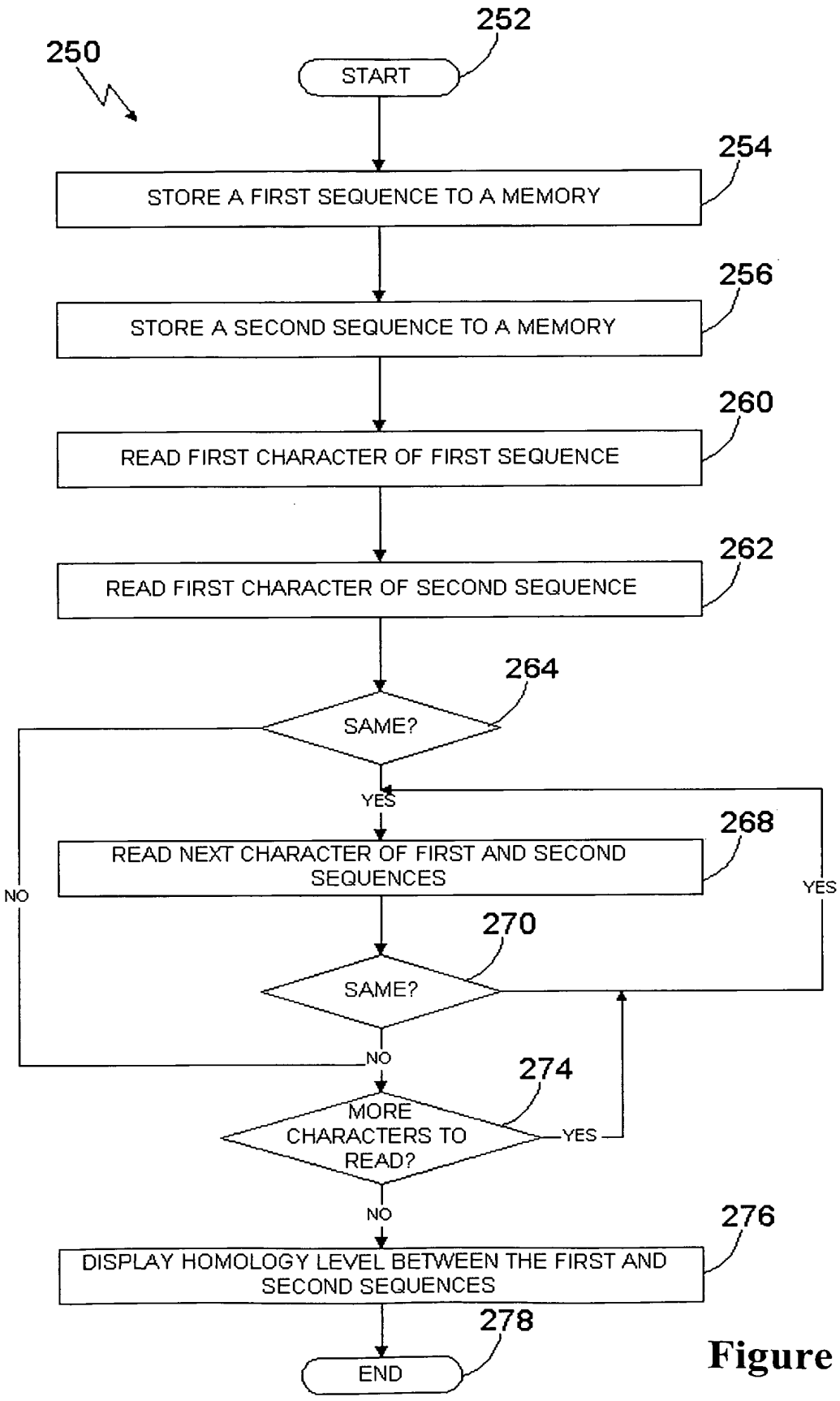
250
252
START
254
STORE A FIRST SEQUENCE TO A MEMORY
256
STORE A SECOND SEQUENCE TO A MEMORY
260
READ FIRST CHARACTER OF FIRST SEQUENCE
262
READ FIRST CHARACTER OF SECOND SEQUENCE
264
SAME?
YES
NO
268
READ NEXT CHARACTER OF FIRST AND SECOND SEQUENCES
YES
270
SAME?
NO
274
MORE CHARACTERS TO READ?
YES
NO
276
DISPLAY HOMOLOGY LEVEL BETWEEN THE FIRST AND SECOND SEQUENCES
278
END

Figure 4

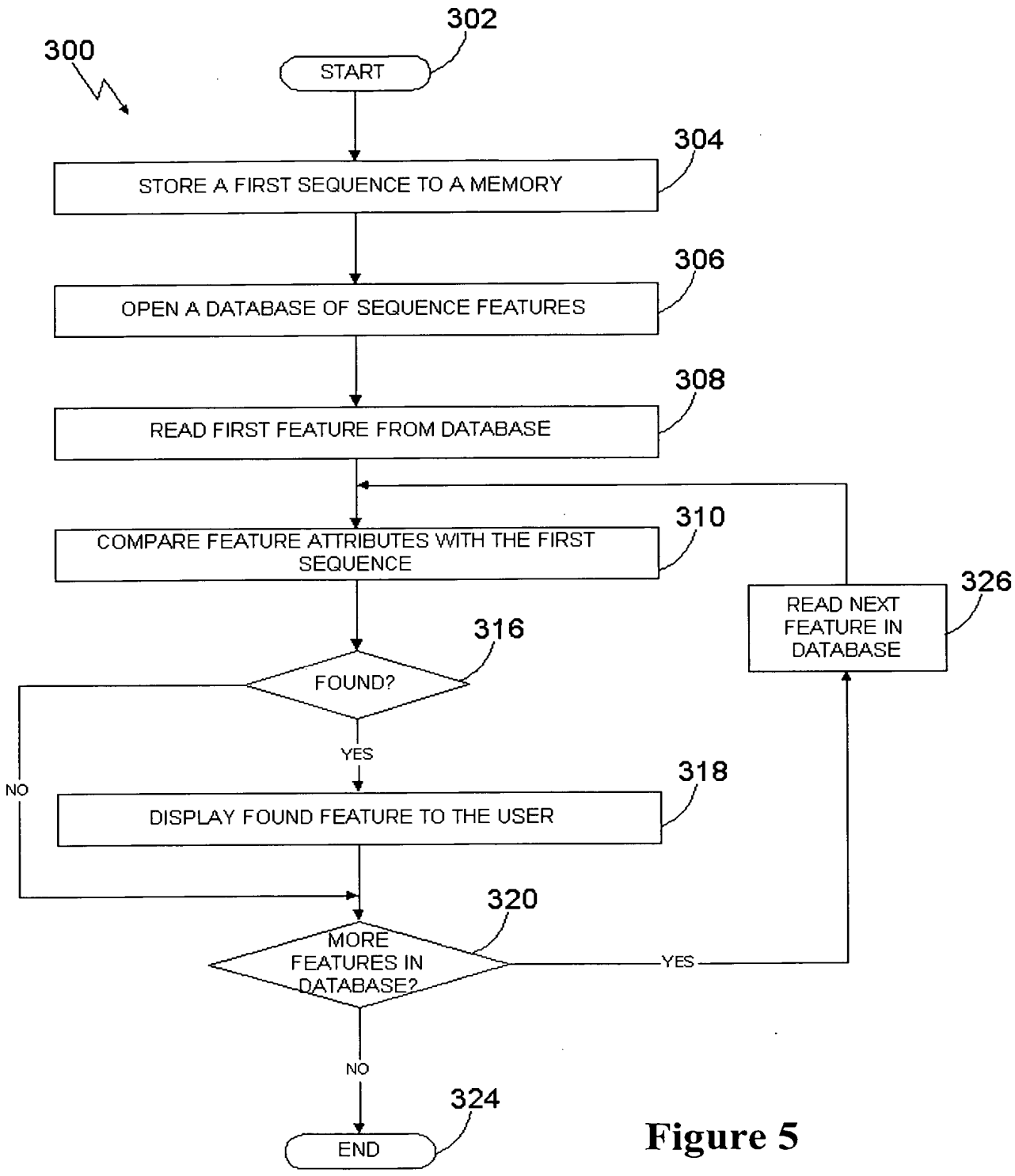
300
302
START
304
STORE A FIRST SEQUENCE TO A MEMORY
306
OPEN A DATABASE OF SEQUENCE FEATURES
308
READ FIRST FEATURE FROM DATABASE
310
COMPARE FEATURE ATTRIBUTES WITH THE FIRST SEQUENCE
316
FOUND?
YES
NO
318
DISPLAY FOUND FEATURE TO THE USER
320
MORE FEATURES IN DATABASE?
YES
NO
324
END
326
READ NEXT FEATURE IN DATABASE

Figure 5

<110> GENSET

<120> Full-length human cDNAs encoding potentially secreted proteins

<130> 78.WO1

<150> US 60/169,629
<151> 1999-12-08

<150> US 60/187,470
<151> 2000-03-06

<160> 482

<170> Patent.pm

<210> 1
<211> 2201
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 169..1692

<220>
<221> sig_peptide
<222> 169..249
<223> Von Heijne matrix
score 7.15265901862021
seq VLLLLLLERGMFS/SP

<400> 1

```
agatgtgaat agctccacta taccagcctc gtcttccttc cgggggacaa cgtgggtcag      60
ggcacagaga gatatttaat gtcaccctct tggggctttc atgggactcc ctctgccaca     120
ttttttggag gttgggaaag ttgctagagg cttcagaact ccagccta atg gat ccc      177
                                                    Met Asp Pro
                                                              -25
aaa ctc ggg aga atg gct gcg tcc ctg ctg gct gtg ctg ctg ctg ctg       225
Lys Leu Gly Arg Met Ala Ala Ser Leu Leu Ala Val Leu Leu Leu Leu
                -20                 -15                 -10
ctg ctg gag cgc ggc atg ttc tcc tca ccc tcc ccg ccc ccg gcg ctg       273
Leu Leu Glu Arg Gly Met Phe Ser Ser Pro Ser Pro Pro Pro Ala Leu
            -5                  1               5
tta gag aaa gtc ttc cag tac att gac ctc cat cag gat gaa ttt gtg       321
Leu Glu Lys Val Phe Gln Tyr Ile Asp Leu His Gln Asp Glu Phe Val
    10                  15                  20
cag acg ctg aag gag tgg gtg gcc atc gag agc gac tct gtc cag cct       369
Gln Thr Leu Lys Glu Trp Val Ala Ile Glu Ser Asp Ser Val Gln Pro
25                  30                  35                  40
gtg cct cgc ttc aga caa gag ctc ttc aga atg atg gcc gtg gct gcg       417
Val Pro Arg Phe Arg Gln Glu Leu Phe Arg Met Met Ala Val Ala Ala
                45                  50                  55
gac acg ctg cag cgc ctg ggg gcc cgt gtg gcc tcg gtg gac atg ggt       465
Asp Thr Leu Gln Arg Leu Gly Ala Arg Val Ala Ser Val Asp Met Gly
            60                  65                  70
cct cag cag ctg ccc gat ggt cag agt ctt cca ata cct ccc gtc atc       513
Pro Gln Gln Leu Pro Asp Gly Gln Ser Leu Pro Ile Pro Pro Val Ile
        75                  80                  85
ctg gcc gaa ctg ggg agc gat ccc acg aaa ggc acc gtg tgc ttc tac       561
Leu Ala Glu Leu Gly Ser Asp Pro Thr Lys Gly Thr Val Cys Phe Tyr
    90                  95                  100
ggc cac ttg gac gtg cag cct gct gac cgg ggc gat ggg tgg ctc acg       609
```

```
Gly His Leu Asp Val Gln Pro Ala Asp Arg Gly Asp Gly Trp Leu Thr
105                 110                 115                 120
gac ccc tat gtg ctg acg gag gta gac ggg aaa ctt tat gga cga gga      657
Asp Pro Tyr Val Leu Thr Glu Val Asp Gly Lys Leu Tyr Gly Arg Gly
                125                 130                 135
gcg acc gac aac aaa ggc cct gtc ttg gct tgg atc aat gct gtg agc      705
Ala Thr Asp Asn Lys Gly Pro Val Leu Ala Trp Ile Asn Ala Val Ser
            140                 145                 150
gcc ttc aga gcc ctg gag caa gat ctt cct gtg aat atc aaa ttc atc      753
Ala Phe Arg Ala Leu Glu Gln Asp Leu Pro Val Asn Ile Lys Phe Ile
        155                 160                 165
att gag ggg atg gaa gag gct ggc tct gtt gcc ctg gag gaa ctt gtg      801
Ile Glu Gly Met Glu Glu Ala Gly Ser Val Ala Leu Glu Glu Leu Val
    170                 175                 180
gaa aaa gaa aag gac cga ttc ttc tct ggt gtg gac tac att gta att      849
Glu Lys Glu Lys Asp Arg Phe Phe Ser Gly Val Asp Tyr Ile Val Ile
185                 190                 195                 200
tca gat aac ctg tgg atc agc caa agg aag cca gca atc act tat gga      897
Ser Asp Asn Leu Trp Ile Ser Gln Arg Lys Pro Ala Ile Thr Tyr Gly
                205                 210                 215
acc cgg ggg aac agc tac ttc atg gtg gag gtg aaa tgc aga gac cag      945
Thr Arg Gly Asn Ser Tyr Phe Met Val Glu Val Lys Cys Arg Asp Gln
            220                 225                 230
gat ttt cac tca gga acc ttt ggt ggc atc ctt cat gaa cca atg gct      993
Asp Phe His Ser Gly Thr Phe Gly Gly Ile Leu His Glu Pro Met Ala
        235                 240                 245
gat ctg gtt gct ctt ctc ggt agc ctg gta gac tcg tct ggt cat atc     1041
Asp Leu Val Ala Leu Leu Gly Ser Leu Val Asp Ser Ser Gly His Ile
    250                 255                 260
ctg gtc cct gga atc tat gat gaa gtg gtt cct ctt aca gaa gag gaa     1089
Leu Val Pro Gly Ile Tyr Asp Glu Val Val Pro Leu Thr Glu Glu Glu
265                 270                 275                 280
ata aat aca tac aaa gcc atc cat cta gac cta gaa gaa tac cgg aat     1137
Ile Asn Thr Tyr Lys Ala Ile His Leu Asp Leu Glu Glu Tyr Arg Asn
                285                 290                 295
agc agc cgg gtt gag aaa ttt ctg ttc gat act aag gag gag att cta     1185
Ser Ser Arg Val Glu Lys Phe Leu Phe Asp Thr Lys Glu Glu Ile Leu
            300                 305                 310
atg cac ctc tgg agg tac cca tct ctt tct att cat ggg atc gag ggc     1233
Met His Leu Trp Arg Tyr Pro Ser Leu Ser Ile His Gly Ile Glu Gly
        315                 320                 325
gcg ttt gat gag cct gga act aaa aca gtc ata cct ggc cga gtt ata     1281
Ala Phe Asp Glu Pro Gly Thr Lys Thr Val Ile Pro Gly Arg Val Ile
    330                 335                 340
gga aaa ttt tca atc cgt cta gtc cct cac atg aat gtg tct gcg gtg     1329
Gly Lys Phe Ser Ile Arg Leu Val Pro His Met Asn Val Ser Ala Val
345                 350                 355                 360
gaa aaa cag gtg aca cga cat ctt gaa gat gtg ttc tcc aaa aga aat     1377
Glu Lys Gln Val Thr Arg His Leu Glu Asp Val Phe Ser Lys Arg Asn
                365                 370                 375
agt tcc aac aag atg gtt gtt tcc atg act cta gga cta cac ccg tgg     1425
Ser Ser Asn Lys Met Val Val Ser Met Thr Leu Gly Leu His Pro Trp
            380                 385                 390
att gca aat att gat gac acc cag tat ctc gca gca aaa aga gcg atc     1473
Ile Ala Asn Ile Asp Asp Thr Gln Tyr Leu Ala Ala Lys Arg Ala Ile
        395                 400                 405
aga aca gtg ttt gga aca gaa cca gat atg atc cgg gat gga tcc acc     1521
Arg Thr Val Phe Gly Thr Glu Pro Asp Met Ile Arg Asp Gly Ser Thr
    410                 415                 420
att cca att gcc aaa atg ttc cag gag atc gtc cac aag agc gtg gtg     1569
Ile Pro Ile Ala Lys Met Phe Gln Glu Ile Val His Lys Ser Val Val
425                 430                 435                 440
cta att ccg ctg gga gct gtt gat gat gga gaa cat tcg cag aat gag     1617
```

```
Leu Ile Pro Leu Gly Ala Val Asp Asp Gly Glu His Ser Gln Asn Glu
                445                 450                 455
aaa atc aac agg tgg aac tac ata gag gga acc aaa tta ttt gct gcc     1665
Lys Ile Asn Arg Trp Asn Tyr Ile Glu Gly Thr Lys Leu Phe Ala Ala
            460                 465                 470
ttt ttc tta gag atg gcc cag ctc cat taatcacaag aaccttctag           1712
Phe Phe Leu Glu Met Ala Gln Leu His
        475                 480
tctgatctga tccactgaca gattcacctc ccccacatcc ctagacaggg atggaatgta   1772
aatatccaga gaatttgggt ctagtatagt acattttccc ttccatttaa aatgtcttgg   1832
gatatctgga tcagtaataa aatatttcaa aggcacagat gttggaaatg gtttaaggtc   1892
ccccactgca caccttcctc aagtcatagc tgcttgcagc aacttgattt ccccaagtcc   1952
tgtgcaatag ccccaggatt ggattccttc caacctttta gcatatctcc aaccttgcaa   2012
tttgattggc ataatcactc cagtttgctt tctaggtcct caagtgctcg tgacacataa   2072
tcattccatc caatgatcgc ctttgcttta ccactctttc cttttatctt attaataaaa   2132
atgttggtct ccaccactga aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaagaaaaaa   2192
aaaaaaaaa                                                          2201

<210> 2
<211> 1631
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 148..1140

<220>
<221> sig_peptide
<222> 148..240
<223> Von Heijne matrix
      score 10.0910253445132
      seq LVLLLVTRSPVNA/CL

<400> 2
gtctgctgcc gccattgtgc ggcgctggtc ccctcagagg gttcctgctg ctgccggtgc     60
cttggaccct ccccctcgct tctcgttcta ctgccccagg agcccggcgg gtccgggact    120
cccgtccgtg ccggtgcggg cgccggc atg tgg ctg tgg gag gac cag ggc ggc    174
                              Met Trp Leu Trp Glu Asp Gln Gly Gly
                                  -30                 -25
ctc ctg ggc cct ttc tcc ttc ctg ctg cta gtg ctg ctg ctg gtg acg      222
Leu Leu Gly Pro Phe Ser Phe Leu Leu Leu Val Leu Leu Leu Val Thr
        -20                 -15                 -10
cgg agc ccg gtc aat gcc tgc ctc ctc acc ggc agc ctc ttc gtt cta      270
Arg Ser Pro Val Asn Ala Cys Leu Leu Thr Gly Ser Leu Phe Val Leu
    -5                  1                   5                   10
ctg cgc gtc ttc agc ttt gag ccg gtg ccc tct tgc agg gcc ctg cag      318
Leu Arg Val Phe Ser Phe Glu Pro Val Pro Ser Cys Arg Ala Leu Gln
                15                  20                  25
gtg ctc aag ccc cgg gac cgc att tct gcc atc gcc cac cgt ggc ggc      366
Val Leu Lys Pro Arg Asp Arg Ile Ser Ala Ile Ala His Arg Gly Gly
            30                  35                  40
agc cac gac gcg ccc gag aac acg ctg gcg gcc att cgg cag gca gct      414
Ser His Asp Ala Pro Glu Asn Thr Leu Ala Ala Ile Arg Gln Ala Ala
        45                  50                  55
aag aat gga gca aca ggc gtg gag ttg gac att gag ttt act tct gac      462
Lys Asn Gly Ala Thr Gly Val Glu Leu Asp Ile Glu Phe Thr Ser Asp
    60                  65                  70
ggg att cct gtc tta atg cac gat aac aca gta gat agg acg act gat      510
Gly Ile Pro Val Leu Met His Asp Asn Thr Val Asp Arg Thr Thr Asp
75                  80                  85                  90
ggg act ggg cga ttg tgt gat ttg aca ttt gaa caa att agg aag ctg      558
Gly Thr Gly Arg Leu Cys Asp Leu Thr Phe Glu Gln Ile Arg Lys Leu
```

```
                  95                  100                 105
aat cct gca gca aac cac aga ctc agg aat gat ttc cct gat gaa aag      606
Asn Pro Ala Ala Asn His Arg Leu Arg Asn Asp Phe Pro Asp Glu Lys
              110                 115                 120
atc cct acc cta atg gaa gct gtt gca gag tgc cta aac cat aac ctc      654
Ile Pro Thr Leu Met Glu Ala Val Ala Glu Cys Leu Asn His Asn Leu
          125                 130                 135
aca atc ttc ttt gat gtc aaa ggc cat gca cac aag gct act gag gct      702
Thr Ile Phe Phe Asp Val Lys Gly His Ala His Lys Ala Thr Glu Ala
      140                 145                 150
cta aag aaa atg tat atg gaa ttt cct caa ctg tat aat aat agt gtg      750
Leu Lys Lys Met Tyr Met Glu Phe Pro Gln Leu Tyr Asn Asn Ser Val
155                 160                 165                 170
gtc tgt tct ttc ttg cca gaa gtt atc tac aag atg aga caa aca gat      798
Val Cys Ser Phe Leu Pro Glu Val Ile Tyr Lys Met Arg Gln Thr Asp
                  175                 180                 185
cgg gat gta ata aca gca tta act cac aga cct tgg agc cta agc cat      846
Arg Asp Val Ile Thr Ala Leu Thr His Arg Pro Trp Ser Leu Ser His
              190                 195                 200
aca gga gat ggg aaa cca cgc tat gat act ttc tgg aaa cat ttt ata      894
Thr Gly Asp Gly Lys Pro Arg Tyr Asp Thr Phe Trp Lys His Phe Ile
          205                 210                 215
ttt gtt atg atg gac att ttg ctc gat tgg agc atg cat aat atc ttg      942
Phe Val Met Met Asp Ile Leu Leu Asp Trp Ser Met His Asn Ile Leu
      220                 225                 230
tgg tac ctg tgt gga att tca gct ttc ctc atg caa aag gat ttt gta      990
Trp Tyr Leu Cys Gly Ile Ser Ala Phe Leu Met Gln Lys Asp Phe Val
235                 240                 245                 250
tcc ccg gcc tac ttg aag aag tgg tca gct aaa gga atc cag gtt gtt     1038
Ser Pro Ala Tyr Leu Lys Lys Trp Ser Ala Lys Gly Ile Gln Val Val
                  255                 260                 265
ggt tgg act gtt aat acc ttt gat gaa aag agt tac tac gaa tcc cat     1086
Gly Trp Thr Val Asn Thr Phe Asp Glu Lys Ser Tyr Tyr Glu Ser His
              270                 275                 280
ctt ggt tcc agc tat atc act gac agc atg gta gaa gac tgc gaa cct     1134
Leu Gly Ser Ser Tyr Ile Thr Asp Ser Met Val Glu Asp Cys Glu Pro
          285                 290                 295
cac ttc tagactttca cggtgggacg aaacgggttc agaaactgcc aggggcctca     1190
His Phe
      300
tacagggata tcaaaatacc ctttgtgcta gcccaggccc tggggaatca ggtgactcac   1250
acaaatgcaa tagttggtca ctgcattttt acctgaacca aagctaaacc cggtgttgcc   1310
accatgcacc atggcatgcc agagttcaac actgttgctc ttgaaaatct ggggtctgaa   1370
aaaacgcaca agagcccctg ccctgcccta gctgaggcac acagggagac ccagtgagga   1430
taagcacaga ttgaattgta caatttgcag atgcagatgt aaatgcatgg gacatgcatg   1490
ataactcaga gttgacattt taaaacttgc cacacttatt tcaaatattt gtactcagct   1550
atgttaacat gtactgtaga catcaaactt gtggccatac taataaaatt attaaaagga   1610
gcacaaaaaa aaaaaaaaaa a                                             1631

<210> 3
<211> 1245
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 85..906

<220>
<221> sig_peptide
<222> 85..135
<223> Von Heijne matrix
      score 3.86022363031904
```

seq GFVAALVAGGVAG/VS

<400> 3

```
aaaacatggc ggcgcccagc gcgcgaggac gtgatccgct tctgctccgg cttggattgt      60
agccttgacg aggtctgagc gacc atg gac cgg ccg ggg ttc gtg gca gcg         111
                            Met Asp Arg Pro Gly Phe Val Ala Ala
                                       -15                 -10
ctg gtg gct ggt ggg gta gca ggt gtt tct gtt gac ttg ata tta ttt       159
Leu Val Ala Gly Gly Val Ala Gly Val Ser Val Asp Leu Ile Leu Phe
            -5                  1                   5
cct ctg gat acc att aaa acc agg ctg cag agt ccc caa gga ttt agt       207
Pro Leu Asp Thr Ile Lys Thr Arg Leu Gln Ser Pro Gln Gly Phe Ser
    10                  15                  20
aag gct ggt ggt ttt cat gga ata tat gct ggc gtt cct tct gct gct       255
Lys Ala Gly Gly Phe His Gly Ile Tyr Ala Gly Val Pro Ser Ala Ala
25                  30                  35                  40
att gga tcc ttt cct aat gct gct gca ttt ttt atc acc tat gaa tat       303
Ile Gly Ser Phe Pro Asn Ala Ala Ala Phe Phe Ile Thr Tyr Glu Tyr
                45                  50                  55
gtg aag tgg ttt ttg cat gct gat tca tct tca tat ttg aca cct atg       351
Val Lys Trp Phe Leu His Ala Asp Ser Ser Ser Tyr Leu Thr Pro Met
            60                  65                  70
aaa cat atg ttg gct gcc tct gct gga gaa gtg gtt gcc tgc ctg att       399
Lys His Met Leu Ala Ala Ser Ala Gly Glu Val Val Ala Cys Leu Ile
        75                  80                  85
cga gtt cca tct gaa gtg gtt aag cag agg gca cag gta tct gct tct       447
Arg Val Pro Ser Glu Val Val Lys Gln Arg Ala Gln Val Ser Ala Ser
    90                  95                  100
aca aga aca ttt cag att ttc tct aac atc tta tat gaa gag ggt atc       495
Thr Arg Thr Phe Gln Ile Phe Ser Asn Ile Leu Tyr Glu Glu Gly Ile
105                 110                 115                 120
caa ggg ttg tat cga ggc tat aaa agc aca gtt tta aga gag att cct       543
Gln Gly Leu Tyr Arg Gly Tyr Lys Ser Thr Val Leu Arg Glu Ile Pro
                125                 130                 135
ttt tct ttg gtc cag ttt ccc tta tgg gag tcc tta aaa gcc ctc tgg       591
Phe Ser Leu Val Gln Phe Pro Leu Trp Glu Ser Leu Lys Ala Leu Trp
            140                 145                 150
tcc tgg agg cag gat cat gtg gtg gat tct tgg cag tca gca gtc tgt       639
Ser Trp Arg Gln Asp His Val Val Asp Ser Trp Gln Ser Ala Val Cys
        155                 160                 165
gga gct ttt gca ggt gga ttt gcc gct gca gtc acc acc cct cta gac       687
Gly Ala Phe Ala Gly Gly Phe Ala Ala Ala Val Thr Thr Pro Leu Asp
    170                 175                 180
gtg gca aag aca aga att atg ctg gca aag gct ggc tcc agc act gct       735
Val Ala Lys Thr Arg Ile Met Leu Ala Lys Ala Gly Ser Ser Thr Ala
185                 190                 195                 200
gat ggg aat gtg ctc tct gtc ctg cat ggg gtc tgg cgg tca cag ggg       783
Asp Gly Asn Val Leu Ser Val Leu His Gly Val Trp Arg Ser Gln Gly
                205                 210                 215
ctg gca gga tta ttt gca ggt gtc ttc cct cga atg gca gcc atc agt       831
Leu Ala Gly Leu Phe Ala Gly Val Phe Pro Arg Met Ala Ala Ile Ser
            220                 225                 230
ctg gga ggt ttc atc ttt ctg ggg gct tat gac cga acg cac agc ttg       879
Leu Gly Gly Phe Ile Phe Leu Gly Ala Tyr Asp Arg Thr His Ser Leu
        235                 240                 245
ctg ttg gaa gtt ggc aga aag agt cct tgaagcagag acaagcctca              926
Leu Leu Glu Val Gly Arg Lys Ser Pro
    250                 255
cctccacttc tgtcaagaga ggggcctgca gtgcaaaccc tcttccgctg agcagctgtc      986
tgaactatag gccccagtgc tgaagaccag ttgtgctaag ataccggcat ggagattgtg     1046
ccatccgtgg tataggctgg ctggtatgaa gtcattggcc tgtatgccag agagctaaga     1106
gaagaaaacg gggtctgtgg cggtactctg aacaatttcc tcagaacctc ttaataaata     1166
agtttggtaa tgctgagaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa     1226
```

agaaaaaaaa aaaaaaaaa 1245

<210> 4
<211> 1623
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 31..1248

<220>
<221> sig_peptide
<222> 31..135
<223> Von Heijne matrix
score 6.3770152988307
seq TLLLFAAPFGLLG/EK

<400> 4

```
aacctcttcc gtcggctgaa ttgcggccgt atg cgc ggc tct gtg gag tgc acc      54
                                  Met Arg Gly Ser Val Glu Cys Thr
                                  -35                 -30
tgg ggt tgg ggg cac tgt gcc ccc agc ccc ctg ctc ctt tgg act cta      102
Trp Gly Trp Gly His Cys Ala Pro Ser Pro Leu Leu Leu Trp Thr Leu
        -25                 -20                 -15
ctt ctg ttt gca gcc cca ttt ggc ctg ctg ggg gag aag acc cgc cag      150
Leu Leu Phe Ala Ala Pro Phe Gly Leu Leu Gly Glu Lys Thr Arg Gln
    -10                 -5                  1                   5
gtg tct ctg gag gtc atc cct aac tgg ctg ggc ccc ctg cag aac ctg      198
Val Ser Leu Glu Val Ile Pro Asn Trp Leu Gly Pro Leu Gln Asn Leu
                10                  15                  20
ctt cat ata cgg gca gtg ggc acc aat tcc aca ctg cac tat gtg tgg      246
Leu His Ile Arg Ala Val Gly Thr Asn Ser Thr Leu His Tyr Val Trp
            25                  30                  35
agc agc ctg ggg cct ctg gca gtg gta atg gtg gcc acc aac acc ccc      294
Ser Ser Leu Gly Pro Leu Ala Val Val Met Val Ala Thr Asn Thr Pro
        40                  45                  50
cac agc acc ctg agc gtc aac tgg agc ctc ctg cta tcc cct gag ccc      342
His Ser Thr Leu Ser Val Asn Trp Ser Leu Leu Leu Ser Pro Glu Pro
    55                  60                  65
gat ggg ggc ctg atg gtg ctc cct aag gac agc att cag ttt tct tct      390
Asp Gly Gly Leu Met Val Leu Pro Lys Asp Ser Ile Gln Phe Ser Ser
70                  75                  80                  85
gcc ctt gtt ttt acc agg ctg ctt gag ttt gac agc acc aac gtg tcc      438
Ala Leu Val Phe Thr Arg Leu Leu Glu Phe Asp Ser Thr Asn Val Ser
                90                  95                  100
gat acg gca gca aag cct ttg gga aga cca tat cct cca tac tcc ttg      486
Asp Thr Ala Ala Lys Pro Leu Gly Arg Pro Tyr Pro Pro Tyr Ser Leu
            105                 110                 115
gcc gat ttc tct tgg aac aac atc act gat tca ttg gat cct gcc acc      534
Ala Asp Phe Ser Trp Asn Asn Ile Thr Asp Ser Leu Asp Pro Ala Thr
        120                 125                 130
ctg agt gcc aca ttt caa ggc cac ccc atg aac gac cct acc agg act      582
Leu Ser Ala Thr Phe Gln Gly His Pro Met Asn Asp Pro Thr Arg Thr
    135                 140                 145
ttt gcc aat ggc agc ctg gcc ttc agg gtc cag gcc ttt tcc agg tcc      630
Phe Ala Asn Gly Ser Leu Ala Phe Arg Val Gln Ala Phe Ser Arg Ser
150                 155                 160                 165
agc cga cca gcc caa ccc cct cgc ctc ctg cac aca gca gac acc tgt      678
Ser Arg Pro Ala Gln Pro Pro Arg Leu Leu His Thr Ala Asp Thr Cys
                170                 175                 180
cag cta gag gtg gcc ctg att gga gcc tct ccc cgg gga aac cgt tcc      726
Gln Leu Glu Val Ala Leu Ile Gly Ala Ser Pro Arg Gly Asn Arg Ser
```

```
                185                 190                 195
ctg ttt ggg ctg gag gta gcc aca ttg ggc cag ggc cct gac tgc ccc      774
Leu Phe Gly Leu Glu Val Ala Thr Leu Gly Gln Gly Pro Asp Cys Pro
        200                 205                 210
tca atg cag gag cag cac tcc atc gac gat gaa tat gca ccg gcc gtc      822
Ser Met Gln Glu Gln His Ser Ile Asp Asp Glu Tyr Ala Pro Ala Val
    215                 220                 225
ttc cag ttg gac cag cta ctg tgg ggc tcc ctc cca tca ggc ttt gca      870
Phe Gln Leu Asp Gln Leu Leu Trp Gly Ser Leu Pro Ser Gly Phe Ala
230                 235                 240                 245
cag tgg cga cca gtg gct tac tcc cag aag ccg ggg ggc cga gaa tca      918
Gln Trp Arg Pro Val Ala Tyr Ser Gln Lys Pro Gly Gly Arg Glu Ser
                250                 255                 260
gcc ctg ccc tgc caa gct tcc cct ctt cat cct gcc tta gca tac tct      966
Ala Leu Pro Cys Gln Ala Ser Pro Leu His Pro Ala Leu Ala Tyr Ser
            265                 270                 275
ctt ccc cag tca ccc att gtc cga gcc ttc ttt ggg tcc cag aat aac     1014
Leu Pro Gln Ser Pro Ile Val Arg Ala Phe Phe Gly Ser Gln Asn Asn
        280                 285                 290
ttc tgt gcc ttc aat ctg acg ttc ggg gct tcc aca ggc cct ggc tat     1062
Phe Cys Ala Phe Asn Leu Thr Phe Gly Ala Ser Thr Gly Pro Gly Tyr
    295                 300                 305
tgg gac caa cac tac ctc agc tgg tcg atg ctc ctg ggt gtg ggc ttc     1110
Trp Asp Gln His Tyr Leu Ser Trp Ser Met Leu Leu Gly Val Gly Phe
310                 315                 320                 325
cct cca gtg gac ggc ttg tcc cca cta gtc ctg ggc atc atg gca gtg     1158
Pro Pro Val Asp Gly Leu Ser Pro Leu Val Leu Gly Ile Met Ala Val
                330                 335                 340
gcc ctg ggt gcc cca ggg ctc atg ctg cta ggg ggc ggc ttg gtt ctg     1206
Ala Leu Gly Ala Pro Gly Leu Met Leu Leu Gly Gly Gly Leu Val Leu
            345                 350                 355
ctg ctg cac cac aag aag tac tca gag tac cag tcc ata aat              1248
Leu Leu His His Lys Lys Tyr Ser Glu Tyr Gln Ser Ile Asn
        360                 365                 370
taaggcccgc tctctggagg gaaggacatt actgaacctg tcttgctgtg cctcgaaact    1308
ctggaggttg gagcatcaag ttccagcccc cttcactccc ccatcttgct tttctgtgga    1368
acctcagagg ccagcctcga cttcctggag acccccaggt ggggcttcct tcatactttg    1428
ttgggggact ttggaggcgg gcaggggaca gggctattga taaggtcccc ttggtgttgc    1488
cttcttgcat ctccacacat ttcccttgga tgggacttgc aggcctaaat gagaggcatt    1548
ctgactggtt ggctgccctg gaaggcaaga aaatagattt attttttttt cacagggcaa    1608
aaaaaaaaaa aaaaa                                                    1623

<210> 5
<211> 1454
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 72..143

<220>
<221> sig_peptide
<222> 72..119
<223> Von Heijne matrix
      score 5.68931280801877
      seq LGMLLGLLMAACT/PS

<400> 5
gtgtctgcca ctcggctgcc ggaggccgaa ggtccctgac tatggctccc cagagcctgc     60
cttcatctag g atg gct cct ctg ggc atg ctg ctt ggg ctg ctg atg gcc     110
             Met Ala Pro Leu Gly Met Leu Leu Gly Leu Leu Met Ala
                 -15                 -10                 -5
```

```
gcc tgc aca cct tct gcc tca gtc atc aga acc tgaaggagtt tgccctgacc    163
Ala Cys Thr Pro Ser Ala Ser Val Ile Arg Thr
                1               5
aacccagaga agagcagcac caaagaaacg gagagaaaag aaaccaaagc cgaggaggag    223
ctggatgccg aagtcctgga ggtgttccac ccgacgcatg agtggcaggc ccttcagcca    283
gggcaggctg tccctgcagg atcccacgta cggctgaatc ttcagactgg ggaaagagag    343
gcaaaactcc aatatgagga caagttccga aataatttga aaggcaaaag gctggatatc    403
aacaccaaca cctacacatc tcaggatctc aagagtgcac tggcaaaatt caaggagggg    463
gcagagatgg agagttcaaa ggaagacaag gcaaggcagg ctgaggtaaa gcggctcttc    523
cgccccattg aggaactgaa gaaagacttt gatgagctga atgttgtcat tgagactgac    583
atgcagatca tggtacggct gatcaacaag ttcaatagtt ccagctccag tttggaagag    643
aagattgctg cgctctttga tcttgaatat tatgtccatc agatggacaa tgcgcaggac    703
ctgctttcct ttggtggtct tcaagtggtg atcaatgggc tgaacagcac agagcccctc    763
gtgaaggagt atgctgcgtt tgtgctgggc gctgcctttt ccagcaaccc caaggtccag    823
gtggaggcca tcgaaggggg agccctgcag aagctgctgg tcatcctggc cacggagcag    883
ccgctcactg caaagggagg tgctcaccgt gcgcgtggtc acactgctct acgacctggt    943
cacggagaag atgttcgccg aggaggaggc tgagctgacc caggagatgt ccccagagaa   1003
gctgcagcag tatcgccagg tacacctcct gccakgcctg tgggaacagg gctggtgcga   1063
gatcacggcc cacctcctgg cgctgcccga gcatgatgcc ygtgagaagg tgctgcwgac   1123
actgggcgtc ctcctgacca cctgccggga ccgctaccgt caggaccccc agctcggcag   1183
gacactggcc agcctgcagg ctgagtacca ggtgctggcc agcctggagc tgcaggatgg   1243
tgaggacgag ggctacttcc aggagctgct gggctctgtc aacagcttgc tgaaggagct   1303
gagatgaggc cccacaccag gactggactg ggatgccgct agtgaggctg aggggtgcca   1363
gcgtgggtgg gcttctcagg caggaggaca tcttggcagt gctggcttgg ccattaaatg   1423
gaaacctgaa ggccaaaaaa aaaaaaaaaa a                                  1454

<210> 6
<211> 1639
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 111..1154

<220>
<221> sig_peptide
<222> 111..197
<223> Von Heijne matrix
      score 4.68065944212013
      seq LLGPLMAACFTFC/LS

<400> 6
agacggtcgc cgccgcgttt gcgcaggggg agctggtcgc cgccgcggcc gcctggaatt     60
gtgggagttg tgtctgccac tcggctgccg gaggccgaag gtccctgact atg gct       116
                                                        Met Ala
ccc cag agc ctg cct tca tct agg atg gct cct ctg ggc atg ctg ctt       164
Pro Gln Ser Leu Pro Ser Ser Arg Met Ala Pro Leu Gly Met Leu Leu
        -25                 -20                 -15
ggg ccg ctg atg gcc gcc tgc ttc acc ttc tgc ctc agt cat cag aac       212
Gly Pro Leu Met Ala Ala Cys Phe Thr Phe Cys Leu Ser His Gln Asn
    -10                 -5                  1               5
ctg aag gag ttt gcc ctg acc aac cca gag aag agc agc acc aaa gaa       260
Leu Lys Glu Phe Ala Leu Thr Asn Pro Glu Lys Ser Ser Thr Lys Glu
                10                  15                  20
aca gag aga aaa gaa acc aaa gcc gag gag gag ctg gat gcc gaa gtc       308
Thr Glu Arg Lys Glu Thr Lys Ala Glu Glu Glu Leu Asp Ala Glu Val
            25                  30                  35
ctg gag gtg ttc cac ccg acg cat gag tgg cag gcc ctt cag cca ggg       356
Leu Glu Val Phe His Pro Thr His Glu Trp Gln Ala Leu Gln Pro Gly
        40                  45                  50
cag gct gtc cct gca gga tcc cac gta cgg ctg aat ctt cag act ggg       404
Gln Ala Val Pro Ala Gly Ser His Val Arg Leu Asn Leu Gln Thr Gly
```

```
     55                  60                  65
gaa aga gag gca aaa ctc caa tat gag gac aag ttc cga aat aat ttg      452
Glu Arg Glu Ala Lys Leu Gln Tyr Glu Asp Lys Phe Arg Asn Asn Leu
70                  75                  80                  85
aaa ggc aaa agg ctg gat atc aac acc aac acc tac aca tct cag gat      500
Lys Gly Lys Arg Leu Asp Ile Asn Thr Asn Thr Tyr Thr Ser Gln Asp
                90                  95                  100
ctc aag agt gca ctg gca aaa ttc aag gag ggg gca gag atg gag agt      548
Leu Lys Ser Ala Leu Ala Lys Phe Lys Glu Gly Ala Glu Met Glu Ser
            105                 110                 115
tca aag gaa gac aag gca agg cag gct gag gta aag cgg ctc ttc cgc      596
Ser Lys Glu Asp Lys Ala Arg Gln Ala Glu Val Lys Arg Leu Phe Arg
        120                 125                 130
ccc att gag gaa ctg aag aaa gac ttt gat gag ctg aat gtt gtc att      644
Pro Ile Glu Glu Leu Lys Lys Asp Phe Asp Glu Leu Asn Val Val Ile
    135                 140                 145
gag act gac atg cag atc atg gta cgg ctg atc aac aag ttc aat agt      692
Glu Thr Asp Met Gln Ile Met Val Arg Leu Ile Asn Lys Phe Asn Ser
150                 155                 160                 165
tcc agc tcc agt ttg gaa gag aag att gct gcg ctc ttt gat ctt gaa      740
Ser Ser Ser Ser Leu Glu Glu Lys Ile Ala Ala Leu Phe Asp Leu Glu
                170                 175                 180
tat tat gtc cat cag atg gac aat gcg cag gac ctg ctt tcc ttt ggt      788
Tyr Tyr Val His Gln Met Asp Asn Ala Gln Asp Leu Leu Ser Phe Gly
            185                 190                 195
ggt ctt caa gtg gtg atc aat ggg ctg aac agc aca gag ccc ctc gtg      836
Gly Leu Gln Val Val Ile Asn Gly Leu Asn Ser Thr Glu Pro Leu Val
        200                 205                 210
aag gag tat gct gcg ttt gtg ctg ggc gct gcc ttt tcc agc aac ccc      884
Lys Glu Tyr Ala Ala Phe Val Leu Gly Ala Ala Phe Ser Ser Asn Pro
    215                 220                 225
aag gtc cag gtg gag gcc atc gaa ggg gga gcc ctg cag aag ctg ctg      932
Lys Val Gln Val Glu Ala Ile Glu Gly Gly Ala Leu Gln Lys Leu Leu
230                 235                 240                 245
gtc atc ctg gcc acg gag cag ccg ctc act gca aag aag aag gtc ctg      980
Val Ile Leu Ala Thr Glu Gln Pro Leu Thr Ala Lys Lys Lys Val Leu
                250                 255                 260
ttt gca ctg tgc tcc ctg ctg cgc cac ttc ccc tat gcc cag cgg cag     1028
Phe Ala Leu Cys Ser Leu Leu Arg His Phe Pro Tyr Ala Gln Arg Gln
            265                 270                 275
ttc ctg aag ctc ggg ggg ctg cag gtc ctg agg acc ctg gtg cag gag     1076
Phe Leu Lys Leu Gly Gly Leu Gln Val Leu Arg Thr Leu Val Gln Glu
        280                 285                 290
aag ggc acg gag gtg ctc gcc gtg cgc gtg gtc aca ctg ctc tac gac     1124
Lys Gly Thr Glu Val Leu Ala Val Arg Val Val Thr Leu Leu Tyr Asp
    295                 300                 305
ctg gtc acg gag aag atg ttc gcc gag gag taggctgagc tgacccagga       1174
Leu Val Thr Glu Lys Met Phe Ala Glu Glu
310                 315
gatgtcccca gagaagctgc agcagtatcg ccaggtacac ctcctgccag gcctgtggga   1234
acagggctgg tgcgagatca cggcccacct cctggcgctg cccgagcatg atgcccgtga   1294
gaaggtgctg cagacactgg gcgtcctcct gaccacctgc cgggaccgct accgtcagga   1354
cccccagctc ggcaggacac tggccagcct gcaggctgag taccaggtgc tggccagcct   1414
ggagctgcag gatggtgagg acgagggcta cttccaggag ctgctgggct ctgtcaacag   1474
cttgctgaag gagctgagat gaggccccac accaggactg gactgggatg ccgctagtga   1534
ggctgagggg tgccagcgtg ggtgggcttc tcaggcagga ggacatcttg gcagtgctgg   1594
cttggccatt aaatggaaac ctgaaggcaa aaaaaaaaaa aaaaa                   1639

<210> 7
<211> 1768
<212> DNA
<213> Homo sapiens
```

<220>
<221> CDS
<222> 66..1256

<220>
<221> sig_peptide
<222> 66..173
<223> Von Heijne matrix
score 4.89555877630516
seq LLLLRLNDAALRA/LQ

<400> 7

```
agaggaggtg gcggtggtgg ccctcgcctg tggcccccgt gctgcttgca ctcgaactcg      60
tcgcc atg gag gag ctc cag gag cct ctg aga gga cag ctc cgg ctc tgc     110
      Met Glu Glu Leu Gln Glu Pro Leu Arg Gly Gln Leu Arg Leu Cys
          -35                 -30                 -25
ttc acg caa gct gcc cgg act agc ctc tta ctg ctc agg ctc aac gac      158
Phe Thr Gln Ala Ala Arg Thr Ser Leu Leu Leu Leu Arg Leu Asn Asp
    -20                 -15                 -10
gct gcc ctg cgg gcg ctg caa gag tgt cag cgg caa cag gta cgg ccg      206
Ala Ala Leu Arg Ala Leu Gln Glu Cys Gln Arg Gln Gln Val Arg Pro
-5                  1               5                   10
gtg att gct ttc caa ggc cac cga ggg tat ctg aga ctc cca ggc cct      254
Val Ile Ala Phe Gln Gly His Arg Gly Tyr Leu Arg Leu Pro Gly Pro
            15                  20                  25
ggt tgg tcc tgc ctc ttc tcc ttc ata gtg tcc cag tgt tgt cag gag      302
Gly Trp Ser Cys Leu Phe Ser Phe Ile Val Ser Gln Cys Cys Gln Glu
        30                  35                  40
ggc gct ggt ggt agc ttg gac ctt gtg tgc caa cgc ttc ctc agg tct      350
Gly Ala Gly Gly Ser Leu Asp Leu Val Cys Gln Arg Phe Leu Arg Ser
    45                  50                  55
ggg cct aac agc ctc cac tgc ctg ggc tca ctc agg gag cgc ctc att      398
Gly Pro Asn Ser Leu His Cys Leu Gly Ser Leu Arg Glu Arg Leu Ile
60                  65                  70                  75
att tgg gca gcc atg gat tct atc cca gcc cca tca tca gtt cag gga      446
Ile Trp Ala Ala Met Asp Ser Ile Pro Ala Pro Ser Ser Val Gln Gly
                80                  85                  90
cac aac ctg act gaa gat gcc aga cat cct gag agt tgg cag aac aca      494
His Asn Leu Thr Glu Asp Ala Arg His Pro Glu Ser Trp Gln Asn Thr
            95                  100                 105
gga ggc tat tct gaa gga gat gca gta tca cag cca cag atg gca cta      542
Gly Gly Tyr Ser Glu Gly Asp Ala Val Ser Gln Pro Gln Met Ala Leu
        110                 115                 120
gag gag gtg tca gtg tca gat cca ctg gca agc aac caa gga cag tca      590
Glu Glu Val Ser Val Ser Asp Pro Leu Ala Ser Asn Gln Gly Gln Ser
    125                 130                 135
ctc cca gga tcc tca agg gag cac atg gca cag tgg gaa gtg aga agc      638
Leu Pro Gly Ser Ser Arg Glu His Met Ala Gln Trp Glu Val Arg Ser
140                 145                 150                 155
cag acc cat gtt cca aac aga gaa cct gtt cag gca ctg cct tcc tct      686
Gln Thr His Val Pro Asn Arg Glu Pro Val Gln Ala Leu Pro Ser Ser
                160                 165                 170
gcc agc cgg aaa cgt ctg gac aag aaa cgt tca gtg cct gta gcc act      734
Ala Ser Arg Lys Arg Leu Asp Lys Lys Arg Ser Val Pro Val Ala Thr
            175                 180                 185
gta gaa ctg gaa gaa aag agg ttc aga act ctg cct tta gtg cca agc      782
Val Glu Leu Glu Glu Lys Arg Phe Arg Thr Leu Pro Leu Val Pro Ser
        190                 195                 200
ccc cta caa ggc ctg acc aat cag gat tta caa gag gga gaa gat tgg      830
Pro Leu Gln Gly Leu Thr Asn Gln Asp Leu Gln Glu Gly Glu Asp Trp
    205                 210                 215
gag caa gaa gat gag gac atg gac ccc aga tta gaa cac agt tcc tca      878
Glu Gln Glu Asp Glu Asp Met Asp Pro Arg Leu Glu His Ser Ser Ser
```

```
220                 225                 230                 235
gtt caa gaa gat tct gaa tcc cca agt cct gaa gat ata cca gac tac      926
Val Gln Glu Asp Ser Glu Ser Pro Ser Pro Glu Asp Ile Pro Asp Tyr
                240                 245                 250
ctc ctg caa tac agg gcc atc cac agt gca gaa cag caa cat gcc tat      974
Leu Leu Gln Tyr Arg Ala Ile His Ser Ala Glu Gln Gln His Ala Tyr
            255                 260                 265
gag cag gac ttt gag aca gat tat gct gaa tac cgc atc ctg cat gcc     1022
Glu Gln Asp Phe Glu Thr Asp Tyr Ala Glu Tyr Arg Ile Leu His Ala
        270                 275                 280
cgt gtt ggg act gca agc caa agg ttc ata gag ctg gga gca gag att     1070
Arg Val Gly Thr Ala Ser Gln Arg Phe Ile Glu Leu Gly Ala Glu Ile
    285                 290                 295
aaa aga gtt cgg cga gga act cca gaa tac aag gtc ctg gaa gac aag     1118
Lys Arg Val Arg Arg Gly Thr Pro Glu Tyr Lys Val Leu Glu Asp Lys
300                 305                 310                 315
ata atc cag gaa tat aaa aag ttc agg aag cag tac cca agt tac aga     1166
Ile Ile Gln Glu Tyr Lys Lys Phe Arg Lys Gln Tyr Pro Ser Tyr Arg
                320                 325                 330
gaa gaa aag cgt cgc tgt gag tac ctt cac cag aaa ttg tcc cac att     1214
Glu Glu Lys Arg Arg Cys Glu Tyr Leu His Gln Lys Leu Ser His Ile
            335                 340                 345
aaa ggt ctc atc ctg gag ttt gag gaa aag aac agg ggc agc             1256
Lys Gly Leu Ile Leu Glu Phe Glu Glu Lys Asn Arg Gly Ser
        350                 355                 360
tgaagttatc aagggaattt ttgagcctct gcttagtgaa acacaaagga acaaagcagc   1316
tataaactaa atagaatgca actatctgct tttcttatgc tgaccactgg agtccatggt   1376
ggcaagtaga gagctgctct aggttcttga ggtttggttt tcattattaa tttttagggt   1436
atgggcactg tgcaaagact ccatagctgt gcctaggagt ctaggaaaag tgacagaggc   1496
ttggcttttt tacctttagt tcagccaagt cattttcaag tcctgagaaa tgacatcatc   1556
ttcaggataa aataatgagg acattagaca aaccaaacta agtgaatttt agcctggtag   1616
cctctctaag gaaacagtaa taataacttc tgataagagt taaaagaact tgtagcatac   1676
ctggatataa tgggaaaggg cctgggtgtt acccatgtac tgaaaatgaa cttttaccaa   1736
catggctaaa aaattaaaaa aaaaaaaaaa aa                                 1768

<210> 8
<211> 1510
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 190..1398

<220>
<221> sig_peptide
<222> 190..252
<223> Von Heijne matrix
      score 5.8172934575094
      seq ALLWAQEVGQVLA/GR

<400> 8
acggttgccc tggcagcgcg cgaggctggt gagtcggcag ccctgtggca gccggcgggc     60
tggtttccat ggttgcacga ttaggaacca ccagctgctg catcccatgg ccaggggtgg    120
cgtccaggtg gcagagcagc taggaacgca aggcctgaac ctggggccag acaccctgct    180
ctcccggcc atg gtc aac gac cct cca gta cct gcc tta ctg tgg gcc cag     231
          Met Val Asn Asp Pro Pro Val Pro Ala Leu Leu Trp Ala Gln
              -20                 -15                 -10
gag gtg ggc caa gtc ttg gca ggc cgt gcc cgc agg ctg ctg ctg cag      279
Glu Val Gly Gln Val Leu Ala Gly Arg Ala Arg Arg Leu Leu Leu Gln
        -5                  1               5
ttt ggg gtg ctc ttc tgc acc atc ctc ctt ttg ctc tgg gtg tct gtc      327
Phe Gly Val Leu Phe Cys Thr Ile Leu Leu Leu Leu Trp Val Ser Val
```

```
10                  15                  20                  25
ttc ctc tat ggc tcc ttc tac tat tcc tat atg ccg aca gtc agc cac      375
Phe Leu Tyr Gly Ser Phe Tyr Tyr Ser Tyr Met Pro Thr Val Ser His
                30                  35                  40
ctc agc cct gtg cat ttc tac tac agg acc gac tgt gat tcc tcc acc      423
Leu Ser Pro Val His Phe Tyr Tyr Arg Thr Asp Cys Asp Ser Ser Thr
            45                  50                  55
acc tca ctc tgc tcc ttc cct gtt gcc aat gtc tcg ctg act aag ggt      471
Thr Ser Leu Cys Ser Phe Pro Val Ala Asn Val Ser Leu Thr Lys Gly
        60                  65                  70
gga cgt gat cgg gtg ctg atg tat gga cag ccg tat cgt gtt acc tta      519
Gly Arg Asp Arg Val Leu Met Tyr Gly Gln Pro Tyr Arg Val Thr Leu
    75                  80                  85
gag ctt gag ctg cca gag tcc cct gtg aat caa gat ttg ggc atg ttc      567
Glu Leu Glu Leu Pro Glu Ser Pro Val Asn Gln Asp Leu Gly Met Phe
90                  95                  100                 105
ttg gtc acc att tcc tgc tac acc aga ggt ggc cga atc atc tcc act      615
Leu Val Thr Ile Ser Cys Tyr Thr Arg Gly Gly Arg Ile Ile Ser Thr
                110                 115                 120
tct tcg cgt tcg gtg atg ctg cat tac cgc tca gac ctg ctc cag atg      663
Ser Ser Arg Ser Val Met Leu His Tyr Arg Ser Asp Leu Leu Gln Met
            125                 130                 135
ctg gac aca ctg gtc ttc tct agc ctc ctg cta ttt ggc ttt gca gag      711
Leu Asp Thr Leu Val Phe Ser Ser Leu Leu Leu Phe Gly Phe Ala Glu
        140                 145                 150
cag aag cag ctg ctg gag gtg gaa ctc tac gca gac tat aga gag aac      759
Gln Lys Gln Leu Leu Glu Val Glu Leu Tyr Ala Asp Tyr Arg Glu Asn
    155                 160                 165
tcg gtg agt gag tac gtg ccg acc act gga gcg atc att gag atc cac      807
Ser Val Ser Glu Tyr Val Pro Thr Thr Gly Ala Ile Ile Glu Ile His
170                 175                 180                 185
agc aag cgc atc cag ctg tat gga gcc tac ctc cgc atc cac gcg cac      855
Ser Lys Arg Ile Gln Leu Tyr Gly Ala Tyr Leu Arg Ile His Ala His
                190                 195                 200
ttc act ggg ctc aga tac ctg cta tac aac ttc ccg atg acc tgc gcc      903
Phe Thr Gly Leu Arg Tyr Leu Leu Tyr Asn Phe Pro Met Thr Cys Ala
            205                 210                 215
ttc ata ggt gtt gcc agc aac ttc acc ttc ctc agc gtc atc gtg ctc      951
Phe Ile Gly Val Ala Ser Asn Phe Thr Phe Leu Ser Val Ile Val Leu
        220                 225                 230
ttc agc tac atg cag tgg gtg tgg ggg ggc atc tgg ccc cga cac cgc      999
Phe Ser Tyr Met Gln Trp Val Trp Gly Gly Ile Trp Pro Arg His Arg
    235                 240                 245
ttc tct ttg cag gtt aac atc cga aaa aga gac aat tcc cgg aag gaa     1047
Phe Ser Leu Gln Val Asn Ile Arg Lys Arg Asp Asn Ser Arg Lys Glu
250                 255                 260                 265
gtc caa cga agg atc tct gct cat cag cca ggt gca ggg cct gaa ggc     1095
Val Gln Arg Arg Ile Ser Ala His Gln Pro Gly Ala Gly Pro Glu Gly
                270                 275                 280
cag gag gag tca act ccg caa tca gat gtt aca gag gat ggt gag agc     1143
Gln Glu Glu Ser Thr Pro Gln Ser Asp Val Thr Glu Asp Gly Glu Ser
            285                 290                 295
cct gaa gat ccc tca ggg aca gag ggt cag ctg tcc gag gag gag aaa     1191
Pro Glu Asp Pro Ser Gly Thr Glu Gly Gln Leu Ser Glu Glu Glu Lys
        300                 305                 310
cca gat cag cag ccc ctg agc gga gaa gag gag cta gag cct gag gcc     1239
Pro Asp Gln Gln Pro Leu Ser Gly Glu Glu Glu Leu Glu Pro Glu Ala
    315                 320                 325
agt gat ggt tca ggc tcc tgg gaa gat gca gct ttg ctg acg gag gcc     1287
Ser Asp Gly Ser Gly Ser Trp Glu Asp Ala Ala Leu Leu Thr Glu Ala
330                 335                 340                 345
aac ctg cct gct cct gct cct gct tct gct tct gcc cct gtc cta gag     1335
Asn Leu Pro Ala Pro Ala Pro Ala Ser Ala Ser Ala Pro Val Leu Glu
```

```
                  350                 355                 360
act ctg ggc agc tct gaa cct gct ggg ggt gct ctc cga cag cgc ccc     1383
Thr Leu Gly Ser Ser Glu Pro Ala Gly Gly Ala Leu Arg Gln Arg Pro
            365                 370                 375
acc tgc tct agt tcc tgaagaaaag gggcagactc ctcacattcc agcactttcc     1438
Thr Cys Ser Ser Ser
        380
cacctgactc ctctcccctc gtttttcctt caataaacta ttttgtgtca gctccaaaaa   1498
aaaaaaaaaa aa                                                      1510

<210> 9
<211> 882
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 78..410

<220>
<221> sig_peptide
<222> 78..155
<223> Von Heijne matrix
      score 10.0731536331164
      seq LWLALVSCILTQA/SA

<400> 9
atggctggcc agaggaggaa cgctttgtgt tctcatcgga gctgcatggg aagtctgcat     60
acagcaaagt gacctgc atg cct cac ctt atg gaa agg atg gtg ggc tct       110
                    Met Pro His Leu Met Glu Arg Met Val Gly Ser
                        -25                 -20
ggc ctc ctg tgg ctg gcc ttg gtc tcc tgc att ctg acc cag gca tct      158
Gly Leu Leu Trp Leu Ala Leu Val Ser Cys Ile Leu Thr Gln Ala Ser
-15                 -10                 -5                  1
gca gtg cag cga ggt tat gga aac ccc att gaa gcc agt tcg tat ggg      206
Ala Val Gln Arg Gly Tyr Gly Asn Pro Ile Glu Ala Ser Ser Tyr Gly
            5                   10                  15
ctg gac ctg gac tgc gga gct cct ggc acc cca gag gct cat gtc tgt      254
Leu Asp Leu Asp Cys Gly Ala Pro Gly Thr Pro Glu Ala His Val Cys
        20                  25                  30
ttt gac ccc tgt cag aat tac acc ctc cta gat ttg ggg ccc atc act      302
Phe Asp Pro Cys Gln Asn Tyr Thr Leu Leu Asp Leu Gly Pro Ile Thr
    35                  40                  45
cgg aga ggt gca cag tct ccc ggt gtc atg aat gga acc cct agc act      350
Arg Arg Gly Ala Gln Ser Pro Gly Val Met Asn Gly Thr Pro Ser Thr
50                  55                  60                  65
gca ggg ttc ctg gtg gcc tgg cct atg gtc ctc ctg act gtc ctc ctg      398
Ala Gly Phe Leu Val Ala Trp Pro Met Val Leu Leu Thr Val Leu Leu
                70                  75                  80
gct tgg ctg ttc tgagagctcc gctgagcatc tggccttgaa gtttgtgttc          450
Ala Trp Leu Phe
            85
ttccctctgg caatggctcc cttcagcact tctgctttcc actccaattc acacaggctt    510
ggtattaaca gaatcaaggc caggctaggt taggaaaagg gaagagcttt caccttcttt    570
aaaactctcg gctgggcgca gtggctcatg cctgtaatcc cagcattttg ggaggctgag    630
gcaggtggat cacctgaggt cagcagttca aaatcagcct ggccaaaatg ctgaaactcc    690
gtctctacta aaaatacaaa aattagccag gcatggtgac aggcgcctgt aatcccagct    750
actcgggagg ccaaggcagg agaattgctc gaactcaggg ggtggaggtt gcagtgagtt    810
gagattgtgc cattgcactc cagcctgggc aacagagcaa gactctgtct caggcaaaaa    870
aaaaaaaaaa aa                                                       882

<210> 10
<211> 1849
```

<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 84..299

<220>
<221> sig_peptide
<222> 84..134
<223> Von Heijne matrix
score 3.86022363031904
seq GFVAALVAGGVAG/VS

<400> 10
aaacatggcg gcgcccagcg cgcgaggacg tgatccgctt ctgctccggc ttggattgta 60
gccttgacga ggtctgagcg acc atg gac cgg ccg ggg ttc gtg gca gcg ctg 113
Met Asp Arg Pro Gly Phe Val Ala Ala Leu
-15 -10
gtg gct ggt ggg gta gca ggt gtt tct gtt gac ttg ata tta ttt cct 161
Val Ala Gly Gly Val Ala Gly Val Ser Val Asp Leu Ile Leu Phe Pro
-5 1 5
ctg gat acc att aaa acc agg ctg cag agt ccc caa gga ttt aat aag 209
Leu Asp Thr Ile Lys Thr Arg Leu Gln Ser Pro Gln Gly Phe Asn Lys
10 15 20 25
gct ggt ggt ttt cat gga ata tat gct ggc gtt cct tct gct gct att 257
Ala Gly Gly Phe His Gly Ile Tyr Ala Gly Val Pro Ser Ala Ala Ile
30 35 40
gga tcc ttt cct aat ggt tgc ctg cct gat tcg agt tcc atc 299
Gly Ser Phe Pro Asn Gly Cys Leu Pro Asp Ser Ser Ser Ile
45 50 55
tgaagtggtt aagcagaggg cacaggtatc tgcttctaca agaacatttc agattttctc 359
taacatctta tatgaagagg gtatccaagg gttgtatcga ggctataaaa gcacagtttt 419
aagagagatt ccttttttctt tggtccagtt tcccttatgg gagtccttaa aagccctctg 479
gtcctggagg caggatcatg tggtggattc ttggcagtca gcagtctgtg gagcttttgc 539
aggtggattt gccgctgcag tcaccacccc tctagacgtg gcgaagacaa gaattatgct 599
ggcaaaggct ggctccagca ctgctgatgg gaatgtgctc tctgtcctgc atggggtctg 659
gcggtcacag gggctggcag gattatttgc aggtgtcttc cctcgaatgg cagccatcag 719
tctgggaggt ttcatctttc tgggggctta tgaccgaacg cacagcttgc tgttggaagt 779
tggcagaaag agtccttgaa gcagagacaa gcctcacctc cacttctgtc aagagagggg 839
cctgcagtgt aaaccctctt ccgctgagca gctgtctgaa ctataggccc cagtgctgaa 899
gaccagttgt gctaagatac cggcatggag attgtgccat ccgtggtata ggctggctgg 959
tatgaagtca ttggcctgta tgccagagag ctaagagaag aaaacggggt ctgtggcagt 1019
actctgaaca atttcctcag aacctcttaa taaataagtt tggtaatgct gaggccaggc 1079
cttttagagc tttcatttga tctgtatctg atctttcatt tcctgccacc tgatggtgga 1139
ttcagcagaa ggcaagatgg ttataattct aaaagaatag cttgtttgtt tgtttgtttg 1199
ggaaaaggag acttggggaa gagttgtgta tgtgggtgtt tctcccccta gttaattcct 1259
gttgtgtaag ggtaggcttt gttgaaaaag aaagaaagat tgaactacag gtgcatagca 1319
agcactcttt ctgggtaact aggctgctgg ttttaattac cctcagattt cacccataaa 1379
aacgcacaat tgtattattt tacagagatg tgtccagcgc cccctgtggt gtgtgagaga 1439
aagcagctgc aactcaagtg actaggtggg cccagctggc ttcgtgcagg agggcacggt 1499
gggtgagcca ttctcgccat tctcatgtca gactgaaagg agggcctggg ccagctttga 1559
aaaggcagga tgaaatggaa aggtcaccac acttagggat tttagacctt gactaacaag 1619
ctccaggtgt agaaaaattc aaaacaaaat gtcaggaatc tagcagtgtt gtctgccctg 1679
gagcaaacaa acagtatgtg attttgcttc gcctatttttt ttttttctttt ttgggggaag 1739
ataattaaag gcagaatgac tgcgtttgta aaagaaggac caccaactat actgacattt 1799
ataaatgaac ctttattaaa gacacttcaa tgcaaaaaaa aaaaaaaaaa 1849

<210> 11
<211> 565
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 55..468

<220>
<221> sig_peptide
<222> 55..99
<223> Von Heijne matrix
score 8.96936032049195
seq FTTLLFLAAVAGA/LV

<400> 11
attccccaga ccttctgcag attctgtggt tatactcact cctcatccca aaga atg 57
Met
-15
aaa ttt acc act ctc ctc ttc ttg gca gct gta gca ggg gcc ctg gtc 105
Lys Phe Thr Thr Leu Leu Phe Leu Ala Ala Val Ala Gly Ala Leu Val
-10 -5 1
tat gct gaa gat gcc tcc tct gac tcg acg ggt gct gat cct gcc cag 153
Tyr Ala Glu Asp Ala Ser Ser Asp Ser Thr Gly Ala Asp Pro Ala Gln
5 10 15
gaa gct ggg acc tct aag cct aat gaa gag atc tca ggt cca gca gaa 201
Glu Ala Gly Thr Ser Lys Pro Asn Glu Glu Ile Ser Gly Pro Ala Glu
20 25 30
cca gct tca ccc cca gag aca acc aca aca gcc cag gag act tcg gcg 249
Pro Ala Ser Pro Pro Glu Thr Thr Thr Thr Ala Gln Glu Thr Ser Ala
35 40 45 50
gca gca gtt cag ggg aca gcc aag gtc acc tca agc agg cag gaa cta 297
Ala Ala Val Gln Gly Thr Ala Lys Val Thr Ser Ser Arg Gln Glu Leu
55 60 65
aac ccc ctg aaa tcc ata gtg gag aaa agt atc tta cta aca gaa caa 345
Asn Pro Leu Lys Ser Ile Val Glu Lys Ser Ile Leu Leu Thr Glu Gln
70 75 80
gcc ctt gca aaa gca gga aaa gga atg cac gga ggc gtg cca ggt gga 393
Ala Leu Ala Lys Ala Gly Lys Gly Met His Gly Gly Val Pro Gly Gly
85 90 95
aaa caa ttc atc gaa aat gga agt gaa ttt gca caa aaa tta ctg aag 441
Lys Gln Phe Ile Glu Asn Gly Ser Glu Phe Ala Gln Lys Leu Leu Lys
100 105 110
aaa ttc agt cta tta aaa cca tgg gca tgagaagctg aataatggga 488
Lys Phe Ser Leu Leu Lys Pro Trp Ala
115 120
tcattggact taaagcctta aataccttg tagcccagag ctattaaaac gaaagcatcc 548
aaaaaaaaaa aaaaaaa 565

<210> 12
<211> 1663
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 152..475

<220>
<221> sig_peptide
<222> 152..244
<223> Von Heijne matrix
score 10.0910253445132
seq LVLLLVTRSPVNA/CL

<400> 12
atgtgtctgc tgccgccatt gtgcggcgct ggtcccctca gagggttcct gctgctgccg 60

```
gtgccttgga ccctcccccc cgcttctcgt tctactgccc caggagcccg gcgggtccgg    120
gactcccgtc cgtgccggtg cgggcgccgg c atg tgg ctg tgg gag gac cag       172
                                    Met Trp Leu Trp Glu Asp Gln
                                        -30                 -25
ggc ggc ctc ctg ggc cct ttc tcc ttc ctg ctg cta gtg ctg ctg ctg      220
Gly Gly Leu Leu Gly Pro Phe Ser Phe Leu Leu Leu Val Leu Leu Leu
               -20                 -15                 -10
gtg acg cgg agc ccg gtc aat gcc tgc ctc ctc acc ggc agc ctc ttc      268
Val Thr Arg Ser Pro Val Asn Ala Cys Leu Leu Thr Gly Ser Leu Phe
           -5                  1                   5
gtt cta ctg cgc gtc ttc agc ttt gag ccg gtg ccc tct tgc agg gcc      316
Val Leu Leu Arg Val Phe Ser Phe Glu Pro Val Pro Ser Cys Arg Ala
    10                  15                  20
ctg cag gtg ctc aag ccc cgg gac cgc att tct gcc atc gcc cac cgt      364
Leu Gln Val Leu Lys Pro Arg Asp Arg Ile Ser Ala Ile Ala His Arg
25                  30                  35                  40
ggc ggc agc aam sag gcg ccc gag aac acg ctg gcg gcc att cgg cag      412
Gly Gly Ser Xaa Xaa Ala Pro Glu Asn Thr Leu Ala Ala Ile Arg Gln
                45                  50                  55
cta aga atg gag caa cag gcg tgg agt tgg aca ttg agt tta ctt ctg      460
Leu Arg Met Glu Gln Gln Ala Trp Ser Trp Thr Leu Ser Leu Leu Leu
            60                  65                  70
acg gga ttc ctg tct taatgcacga taacacagta gataggacga ctgatgggac      515
Thr Gly Phe Leu Ser
        75
tgggcgattg tgtgatttga catttgaaca aattaggaag ctgaatcctg cagcaaacca    575
cagactcagg aatgatttcc ctgatgaaaa gatccctacc ctaagggaag ctgttgcaga    635
gtgcctaaac cataacctca caatcttctt tgatgtcaaa ggccatgcac acaaggctac    695
tgaggctcta aagaaaatgt atatggaatt tcctcaactg tataataata gtgtggtctg    755
ttctttcttg ccagaagtta tctacaaggt aacattcggg atttttcttg tacatattag    815
atgagacaaa cagatcggga tgtaataaca gcattaactc acagaccttg gagcctaagc    875
catacaggag atgggaaacc acgctatgat actttctgga aacattttat atttgttatg    935
atggacattt tgctcgattg gagcatgcat aatatcttgt ggtacctgtg tggaatttca    995
gctttcctca tgcaaaagga ttttgtatcc ccggcctact tgaagaagtg gtcagctaaa   1055
ggaatccagg ttgttggttg gactgttaat acctttgatg aaaagagtta ctacgaatcc   1115
catcttggtt ccagctatat cactgacagc atggtagaag actgcgaacc tcacttctag   1175
actttcacgg tgggacgaaa cgggttcaga aactgccagg ggcctcatac agggatatca   1235
aaataccctt tgtgctagcc caggccctgg ggaatcaggt gactcacaca aatgcaatag   1295
ttggtcactg catttttacc tgaaccaaag ctaaacccgg tgttgccacc atgcaccatg   1355
gcatgccaga gttcaacact gttgctcttg aaaatctggg tctgaaaaaa cgcacaagag   1415
cccctgccct gccctagctg aggcacacag ggagacccag tgaggataag cacagattga   1475
attgtacaat ttgcagatgc agatgtaaat gcatgggaca tgcatgataa ctcagagttg   1535
acattttaaa acttgccaca cttatttcaa atatttgtac tcagctatgt taacatgtac   1595
tgtagacatc aaacttgtgg ccatactaat aaaattatta aaaggagcac taaaaaaaaa   1655
aaaaaaaa                                                           1663

<210> 13
<211> 744
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 112..552

<220>
<221> sig_peptide
<222> 112..183
<223> Von Heijne matrix
      score 11.7298925418815
      seq FVLGLGLTPPTLA/QD

<400> 13
```

```
tcacaactgg aacccatctc caggaacaaa cagctggaac ccatctcccg ttgaagggaa      60
actgccagat ttttgtaaga ttcttcctcc tgggagcctg tgttggaaga g atg gtg      117
                                                          Met Val
atg ggc ctg ggc gtt ttg ttg ttg gtc ttc gtg ctg ggt ctg ggt ctg      165
Met Gly Leu Gly Val Leu Leu Leu Val Phe Val Leu Gly Leu Gly Leu
        -20                 -15                 -10
acc cca ccg acc ctg gct cag gat aac tcc agg tac aca cac ttc ctg      213
Thr Pro Pro Thr Leu Ala Gln Asp Asn Ser Arg Tyr Thr His Phe Leu
    -5                  1                   5                   10
acc cag cac tat gat gcc aaa cca cag ggc cgg gat gac aga tac tgt      261
Thr Gln His Tyr Asp Ala Lys Pro Gln Gly Arg Asp Asp Arg Tyr Cys
                15                  20                  25
gaa agc atc atg agg aga cgg ggc ctg acc tca ccc tgc aaa gac atc      309
Glu Ser Ile Met Arg Arg Arg Gly Leu Thr Ser Pro Cys Lys Asp Ile
            30                  35                  40
aac aca ttt att cat ggc aac aag cgc acg atc aag gcc atc tgt gaa      357
Asn Thr Phe Ile His Gly Asn Lys Arg Thr Ile Lys Ala Ile Cys Glu
        45                  50                  55
aac aag aat gga aac cct cac aga gaa aac cta aga ata agc aag tct      405
Asn Lys Asn Gly Asn Pro His Arg Glu Asn Leu Arg Ile Ser Lys Ser
    60                  65                  70
tct ttc cag gtc acc act tgc aag cta cat gga ggt tcc ccc tgg cct      453
Ser Phe Gln Val Thr Thr Cys Lys Leu His Gly Gly Ser Pro Trp Pro
75                  80                  85                  90
cca tgc cag tac cga gcc aca gcg ggg ttc aga aac gtt gtt gtt gct      501
Pro Cys Gln Tyr Arg Ala Thr Ala Gly Phe Arg Asn Val Val Val Ala
                95                  100                 105
tgt gaa aat ggc tta cct gtc cac ttg gat cag tca att ttc cgt cgt      549
Cys Glu Asn Gly Leu Pro Val His Leu Asp Gln Ser Ile Phe Arg Arg
            110                 115                 120
ccg taaccagcgg gcccctggtc aagtgctggc tctgctgtcc ttgccttcca           602
Pro
tttcccctct gcacccagaa cagtggtggc aacattcatt gccaagggcc caaagaaaga    662
gctacctgga ccttttgttt tctgtttgac aacatgttta ataaataaaa atgtcttgat    722
atcagcaaaa aaaaaaaaaa aa                                             744

<210> 14
<211> 1759
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 101..1243

<220>
<221> sig_peptide
<222> 101..199
<223> Von Heijne matrix
      score 3.57142340200611
      seq FLCLGMALCPRQA/TR

<400> 14
gtagagtgct gaaggtcctg ccaacggctc tcttggcgtc tcaacgttcg gatcagcagc      60
ttttttccat tctctctctc cacttcttca gtgagcagcc atg agt tgg act gtg      115
                                              Met Ser Trp Thr Val
                                                          -30
cct gtt gtg cgg gcc agc cag aga gtg agc tcg gtg gga gcg aat ttc      163
Pro Val Val Arg Ala Ser Gln Arg Val Ser Ser Val Gly Ala Asn Phe
            -25                 -20                 -15
cta tgc ctg ggg atg gcc ctg tgt ccg cgt caa gca acg cgc atc ccg      211
Leu Cys Leu Gly Met Ala Leu Cys Pro Arg Gln Ala Thr Arg Ile Pro
        -10                 -5                  1
```

```
ctc aac ggc acc tgg ctc ttc acc ccc gtg agc aag atg gcg act gtg      259
Leu Asn Gly Thr Trp Leu Phe Thr Pro Val Ser Lys Met Ala Thr Val
5                   10                  15                  20
aag agt gag ctt att gag cgt ttc act tcc gag aag ccc gtt cat cac      307
Lys Ser Glu Leu Ile Glu Arg Phe Thr Ser Glu Lys Pro Val His His
                25                  30                  35
agt aag gtc tcc atc ata gga act gga tcg gtg ggc atg gcc tgc gct      355
Ser Lys Val Ser Ile Ile Gly Thr Gly Ser Val Gly Met Ala Cys Ala
            40                  45                  50
atc agc atc tta tta aaa ggc ttg agt gat gaa ctt gcc ctt gtg gat      403
Ile Ser Ile Leu Leu Lys Gly Leu Ser Asp Glu Leu Ala Leu Val Asp
        55                  60                  65
ctt gat gaa gac aaa ctg aag ggt gag acg atg gat ctt caa cat ggc      451
Leu Asp Glu Asp Lys Leu Lys Gly Glu Thr Met Asp Leu Gln His Gly
    70                  75                  80
agc cct ttc acg aaa atg cca aat att gtt tgt agc aaa gat tac ttt      499
Ser Pro Phe Thr Lys Met Pro Asn Ile Val Cys Ser Lys Asp Tyr Phe
85                  90                  95                  100
gtc aca gca aac tcc aac cta gtg att atc aca gca ggt gca cgc caa      547
Val Thr Ala Asn Ser Asn Leu Val Ile Ile Thr Ala Gly Ala Arg Gln
                105                 110                 115
gaa aag gga gaa acg cgc ctt aat tta gtc cag cga aat gtg gcc atc      595
Glu Lys Gly Glu Thr Arg Leu Asn Leu Val Gln Arg Asn Val Ala Ile
            120                 125                 130
ttc aag tta atg att tcc agt att gtc cag tac agc ccc cac tgc aaa      643
Phe Lys Leu Met Ile Ser Ser Ile Val Gln Tyr Ser Pro His Cys Lys
        135                 140                 145
ctg att att gtt tcc aat cca gtg gat atc tta act tat gta gct tgg      691
Leu Ile Ile Val Ser Asn Pro Val Asp Ile Leu Thr Tyr Val Ala Trp
    150                 155                 160
aag ttg agt gca ttt ccc aaa aac cgt att att gga agc ggc tgt aat      739
Lys Leu Ser Ala Phe Pro Lys Asn Arg Ile Ile Gly Ser Gly Cys Asn
165                 170                 175                 180
ctg gat act gct cgt ttt cgt ttc ttg att gga caa aag ctt ggt atc      787
Leu Asp Thr Ala Arg Phe Arg Phe Leu Ile Gly Gln Lys Leu Gly Ile
                185                 190                 195
cat tct gaa agc tgc cat gga tgg atc ctc gga gag cat gga gac tca      835
His Ser Glu Ser Cys His Gly Trp Ile Leu Gly Glu His Gly Asp Ser
            200                 205                 210
agt gtt cct gtg tgg agt gga gtg aac ata gct ggt gtc cct ttg aag      883
Ser Val Pro Val Trp Ser Gly Val Asn Ile Ala Gly Val Pro Leu Lys
        215                 220                 225
gat ctg aac tct gat ata gga act gat aaa gat cct gag caa tgg aaa      931
Asp Leu Asn Ser Asp Ile Gly Thr Asp Lys Asp Pro Glu Gln Trp Lys
    230                 235                 240
aat gtc cac aaa gaa gtg act gca act gcc tat gag att att aaa atg      979
Asn Val His Lys Glu Val Thr Ala Thr Ala Tyr Glu Ile Ile Lys Met
245                 250                 255                 260
aaa ggt tat act tct tgg gcc att ggc cta tct gtg gcc gat tta aca     1027
Lys Gly Tyr Thr Ser Trp Ala Ile Gly Leu Ser Val Ala Asp Leu Thr
                265                 270                 275
gaa agt att ttg aag aat ctt agg aga ata cat cca gtt tcc acc ata     1075
Glu Ser Ile Leu Lys Asn Leu Arg Arg Ile His Pro Val Ser Thr Ile
            280                 285                 290
att aag ggc ctc tat gga ata gat gaa gaa gta ttc ctc agt att cct     1123
Ile Lys Gly Leu Tyr Gly Ile Asp Glu Glu Val Phe Leu Ser Ile Pro
        295                 300                 305
tgt atc ctg gga gag aac ggt att acc aac ctt ata aag ata aag ctg     1171
Cys Ile Leu Gly Glu Asn Gly Ile Thr Asn Leu Ile Lys Ile Lys Leu
    310                 315                 320
acc cct gaa gaa gag gcc cat ctg aaa aaa agt gca aaa aca ctc tgg     1219
Thr Pro Glu Glu Glu Ala His Leu Lys Lys Ser Ala Lys Thr Leu Trp
325                 330                 335                 340
```

```
gaa att cag aat aag ctt aag ctt taaagttgcc taaaactacc attccgaaat    1273
Glu Ile Gln Asn Lys Leu Lys Leu
                345
tattgaagag atcatagata caggattata taacgaaatt ttgaataaac ttgaattcct   1333
aaaagatgga aacaggaaag taggtagagt gattttccta tttatttagt cctccagctc   1393
ttttattgag catccacgtg ctggacgata cttatttaca attcctaagt atttttggta   1453
cctctgatgt agcagcactt gccatgttat atatatgtag ttggcatttg gttcccaaaa   1513
agtaggatgt aggtatttat tgtgttctag aaattccgac tcttttcatt agatatatgc   1573
tatttctttc attcttgctg gtttatacct atgttcattt atatgctgta aaaaagtagt   1633
agcttcttct acaatgtaaa aataaatgta catacaaaaa aatgcagtag tatatacaat   1693
cttttgtttt gcttcctttg atagttaata aattccgttt gttgaatcaa taaaaaaaaa   1753
aaaaaa                                                             1759

<210> 15
<211> 1755
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 101..517

<220>
<221> sig_peptide
<222> 101..199
<223> Von Heijne matrix
      score 3.57613483592743
      seq FLCLGMALCLRQA/TR

<400> 15
gtagagtgct gaaggtcctg ccaacggctc tcttggcgtc tcaacgttcg gatcagcagc     60
ttttttccat tctctctctc cacttcttca gtgagcagcc atg agt tgg act gtg      115
                                              Met Ser Trp Thr Val
                                                             -30
cct gtt gtg cgg gcc agc cag aga atg agc tcg gtg gga gcg aat ttc      163
Pro Val Val Arg Ala Ser Gln Arg Met Ser Ser Val Gly Ala Asn Phe
            -25                 -20                 -15
cta tgc ctg ggg atg gcc ctg tgt ctg cgt caa gca acg cgc atc ccg      211
Leu Cys Leu Gly Met Ala Leu Cys Leu Arg Gln Ala Thr Arg Ile Pro
        -10                 -5                  1
ctc aac ggc acc tgg ctc ttc aca ccc gtg agc aag atg gcg act gtg      259
Leu Asn Gly Thr Trp Leu Phe Thr Pro Val Ser Lys Met Ala Thr Val
5                   10                  15                  20
aag agt gag ctt att gag cgt ttc act tcc gag aag ccc gtt cat cac      307
Lys Ser Glu Leu Ile Glu Arg Phe Thr Ser Glu Lys Pro Val His His
                25                  30                  35
agt aag gtc tcc atc ata gga act gga tcg gtg ggc atg gcc tgc gct      355
Ser Lys Val Ser Ile Ile Gly Thr Gly Ser Val Gly Met Ala Cys Ala
            40                  45                  50
atc agc atc ttg tta aaa ggc ttg agt gat gaa ctt gcc ctt gtg gat      403
Ile Ser Ile Leu Leu Lys Gly Leu Ser Asp Glu Leu Ala Leu Val Asp
        55                  60                  65
ctt gat gaa gac aaa ctg aag ggt gag acg atg gat ctt caa cat ggc      451
Leu Asp Glu Asp Lys Leu Lys Gly Glu Thr Met Asp Leu Gln His Gly
    70                  75                  80
agc cct ttc acg aaa atg cca ata ttg ttt gta gca aag att act ttg      499
Ser Pro Phe Thr Lys Met Pro Ile Leu Phe Val Ala Lys Ile Thr Leu
85                  90                  95                  100
tca cag caa act cca acc tagtgattat cacagcaggt gcacgccaag             547
Ser Gln Gln Thr Pro Thr
                105
aaaagggaga aacgcgcctt aatttagtcc agcgaaatgt ggccatcttc aagtaatgat    607
ttccagtatt gtccagtaca gcccccactg caaactgatt attgtttcca atccagtgga    667
```

```
tatcttaact tatgtagctt ggaagttgag tgcatttccc aaaaaccgta ttattggaag    727
cggctgtaat ctggatactg ctcgttttcg tttcttgatt ggacaaaagc ttggtatcca    787
ttctgaaagc tgccatggat ggatcctcgg agagcatgga gactcaagtg ttcctgtgtg    847
gagtggagtg aacatagctg gtgtcccttt gaaggatctg aactctgata taggaactga    907
taaagatcct gagcaggaaa aatgtccaca aagaagtgac tgcaactgcc tatgagatta    967
ttaaaatgaa aggttatact tcttgggcca ttggcctatc tgtggccgat ttaacagaaa   1027
gtattttgaa gaatcttagg agaatacatc cagtttccac cataactaag ggcctctatg   1087
gaatagatga agaagtattc ctcagtattc cttgtatcct gggagagaac ggtattacca   1147
accttataaa gataaagctg acccctgaag aagaggccca tctgaaaaaa agtgcaaaaa   1207
cactctggga aattcagaat aagcttaagc tttaaagttg cctaaaacta ccattccgaa   1267
attattgaag agatcataga tacaggatta tataacgaaa ttttgaataa acttgaattc   1327
ctaaaagatg gaaacaggaa agtaggtaga gtgattttcc tatttattta gtcctccagc   1387
tcttttattg agcatccacg tgctggacga tacttattta caattcctaa gtatttttgg   1447
tacctctgat gtagcagcac ttgccatgtt atatatatgt agttggcatt tggttcccaa   1507
aaagtaggat gtaggtattt attgtgttct agaaattccg actcttttca ttagatatat   1567
gctatttctt tcattcttgc tggtttatac ctatgttcat ttatatgctg taaaaaagta   1627
gtagcttctt ctacaatgta aaaataaatg tacatacaaa aaaatgcagt agtatataca   1687
atcttttgtt ttgcttcctt tgatagttaa taaattccgt ttgttgaatc aataaaaaaa   1747
aaaaaaaa                                                           1755

<210> 16
<211> 936
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 59..853

<220>
<221> sig_peptide
<222> 59..100
<223> Von Heijne matrix
      score 5.2402423806254
      seq NFILFIFIPGVFS/LK

<400> 16
agaaaggagg ctctgggtag acgcactaga ttactggata aatcacttca atttccca       58
atg aat ttt ata ttg ttt att ttt ata cct gga gtt ttt tcc tta aaa       106
Met Asn Phe Ile Leu Phe Ile Phe Ile Pro Gly Val Phe Ser Leu Lys
                -10                 -5                  1
agt agc act ttg aag cct act att gaa gca ttg cct aat gtg cta cct       154
Ser Ser Thr Leu Lys Pro Thr Ile Glu Ala Leu Pro Asn Val Leu Pro
        5                   10                  15
tta aat gaa gat gtt aat aag cag gaa gaa aag aat gaa gat cat act       202
Leu Asn Glu Asp Val Asn Lys Gln Glu Glu Lys Asn Glu Asp His Thr
    20                  25                  30
ccc aat tat gct cct gct aat gag aaa aat ggc aat tat tat aaa gat       250
Pro Asn Tyr Ala Pro Ala Asn Glu Lys Asn Gly Asn Tyr Tyr Lys Asp
35                  40                  45                  50
ata aaa caa tat gtg ttc aca aca caa aat cca aat ggc act gag tct       298
Ile Lys Gln Tyr Val Phe Thr Thr Gln Asn Pro Asn Gly Thr Glu Ser
                55                  60                  65
gaa ata tct gtg aga gcc aca act gac ctg aat ttt gct cta aaa aac       346
Glu Ile Ser Val Arg Ala Thr Thr Asp Leu Asn Phe Ala Leu Lys Asn
            70                  75                  80
gga tca acc cca aac gtg cct gca ttt tgg aca atg tta gct aaa gct       394
Gly Ser Thr Pro Asn Val Pro Ala Phe Trp Thr Met Leu Ala Lys Ala
        85                  90                  95
ata aat gga aca gca gtg gtc atg gat gat aaa gat caa tta ttt cac       442
Ile Asn Gly Thr Ala Val Val Met Asp Asp Lys Asp Gln Leu Phe His
    100                 105                 110
cca att cca gag tct gat gtg aat gct aca cag gga gaa aat cag cca       490
```

```
Pro Ile Pro Glu Ser Asp Val Asn Ala Thr Gln Gly Glu Asn Gln Pro
115                 120                 125                 130
gat cta gag gat ctg aag atc aaa ata atg ctg gga atc tcg ttg atg      538
Asp Leu Glu Asp Leu Lys Ile Lys Ile Met Leu Gly Ile Ser Leu Met
                135                 140                 145
acc ctc ctc ctc ttt gtg gtc ctc ttg gca ttc tgt agt gct aca ctg      586
Thr Leu Leu Leu Phe Val Val Leu Leu Ala Phe Cys Ser Ala Thr Leu
            150                 155                 160
tac aaa ctg agg cat ctg agt tat aaa agt tgt gag agt cag tac tct      634
Tyr Lys Leu Arg His Leu Ser Tyr Lys Ser Cys Glu Ser Gln Tyr Ser
        165                 170                 175
gtc aac cca gag ctg gcc acg atg tct tac ttt cat cca tca gaa ggt      682
Val Asn Pro Glu Leu Ala Thr Met Ser Tyr Phe His Pro Ser Glu Gly
    180                 185                 190
gtt tca gat aca tcc ttt tcc aag agt gca gag agc agc aca ttt ttg      730
Val Ser Asp Thr Ser Phe Ser Lys Ser Ala Glu Ser Ser Thr Phe Leu
195                 200                 205                 210
ggt acc act tct tca gat atg aga aga tca ggc aca aga aca tca gaa      778
Gly Thr Thr Ser Ser Asp Met Arg Arg Ser Gly Thr Arg Thr Ser Glu
                215                 220                 225
tct aag ata atg acg gat atc att tcc ata ggc tca gat aat gag atg      826
Ser Lys Ile Met Thr Asp Ile Ile Ser Ile Gly Ser Asp Asn Glu Met
            230                 235                 240
cat gaa aac gat gag tcg gtt acc cgg tgaagaaatc aaggaacccg            873
His Glu Asn Asp Glu Ser Val Thr Arg
        245                 250
gtgaagaaat cttattgatg aataaataac tttaattatt ttgtcatcaa aaaaaaaaaa    933
aaa                                                                 936

<210> 17
<211> 747
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 73..672

<220>
<221> sig_peptide
<222> 73..132
<223> Von Heijne matrix
      score 5.21332530399231
      seq SPVFLVFPPEITA/SE

<400> 17
acaagaaaag aacatggtct agactgaagt accaactaaa tcatctcctt tcaaattatc     60
accgacacca tc atg gat tca agc acc gca cac agt ccg gtg ttt ctg gta    111
              Met Asp Ser Ser Thr Ala His Ser Pro Val Phe Leu Val
              -20                 -15                 -10
ttt cct cca gaa atc act gct tca gaa tat gag tcc aca gaa ctt tca      159
Phe Pro Pro Glu Ile Thr Ala Ser Glu Tyr Glu Ser Thr Glu Leu Ser
        -5                  1                   5
gcc acg acc ttt tca act caa agc ccc ttg caa aaa tta ttt gct aga      207
Ala Thr Thr Phe Ser Thr Gln Ser Pro Leu Gln Lys Leu Phe Ala Arg
10                  15                  20                  25
aaa atg aaa atc tta ggg act atc cag atc ctg ttt gga att atg acc      255
Lys Met Lys Ile Leu Gly Thr Ile Gln Ile Leu Phe Gly Ile Met Thr
                30                  35                  40
ttt tct ttt gga gtt atc ttc ctt ttc act ttg tta aaa cca tat cca      303
Phe Ser Phe Gly Val Ile Phe Leu Phe Thr Leu Leu Lys Pro Tyr Pro
            45                  50                  55
agg ttt ccc ttt ata ttt ctt tca gga tat cca ttc tgg ggc tct gtt      351
```

```
Arg Phe Pro Phe Ile Phe Leu Ser Gly Tyr Pro Phe Trp Gly Ser Val
        60                  65                  70
ttg ttc att aat tct gga gcc ttc cta att gca gtg aaa aga aaa acc      399
Leu Phe Ile Asn Ser Gly Ala Phe Leu Ile Ala Val Lys Arg Lys Thr
    75                  80                  85
aca gaa act ctg ata ata ttg agc cga ata atg aat ttt ctt agt gcc      447
Thr Glu Thr Leu Ile Ile Leu Ser Arg Ile Met Asn Phe Leu Ser Ala
90                  95                  100                 105
ctg gga gca ata gct gga atc att ctc ctc aca ttt ggt ttc atc cta      495
Leu Gly Ala Ile Ala Gly Ile Ile Leu Leu Thr Phe Gly Phe Ile Leu
                110                 115                 120
gat caa aac tac att tgt ggt tat tct cac caa aat agt cag tgt aag      543
Asp Gln Asn Tyr Ile Cys Gly Tyr Ser His Gln Asn Ser Gln Cys Lys
            125                 130                 135
gct gtt act gtc ctg ttc ttg gga att ttg att aca ttg atg act ttc      591
Ala Val Thr Val Leu Phe Leu Gly Ile Leu Ile Thr Leu Met Thr Phe
        140                 145                 150
agc att att gaa tta ttc att tct ctg cct ttc tca att ttg ggg tgc      639
Ser Ile Ile Glu Leu Phe Ile Ser Leu Pro Phe Ser Ile Leu Gly Cys
    155                 160                 165
cac tca gag gat tgt gat tgt gaa caa tgt tgt tgactagcac tgtgagaata    692
His Ser Glu Asp Cys Asp Cys Glu Gln Cys Cys
170                 175                 180
aagatgtgtt aaaatattaa aaaaaaaaaa aaaaaaaaag aaaaaaaaaa aaaaa         747
```

<210> 18
<211> 1884
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 94..1275

<220>
<221> sig_peptide
<222> 94..210
<223> Von Heijne matrix
score 4.55778392992629
seq LVLVKRLLAVSVS/CI

<400> 18

```
acagcgcgtg cagcctcgtg cagctcttct ggtctccggc gcccgcccct cagacgtaat     60
gttgaattaa agaaaatact ttatcagaag aag atg gcc act gcc cag ttg cag     114
                                    Met Ala Thr Ala Gln Leu Gln
                                                    -35
agg act ccc atg agt gca ctg gta ttt ccc aat aag ata tca act gaa      162
Arg Thr Pro Met Ser Ala Leu Val Phe Pro Asn Lys Ile Ser Thr Glu
        -30                 -25                 -20
cac cag tct ttg gtg tta gtg aag agg ctt cta gca gtt tca gta tcc      210
His Gln Ser Leu Val Leu Val Lys Arg Leu Leu Ala Val Ser Val Ser
    -15                 -10                 -5
tgt atc acg tat ttg agg gga ata ttc cca gaa tgc gct tat gga aca      258
Cys Ile Thr Tyr Leu Arg Gly Ile Phe Pro Glu Cys Ala Tyr Gly Thr
1               5                   10                  15
aga tat cta gat gat ctt tgt gtc aaa ata ctg aga gaa gat aaa aat      306
Arg Tyr Leu Asp Asp Leu Cys Val Lys Ile Leu Arg Glu Asp Lys Asn
            20                  25                  30
tgc cca gga tct aca cag tta gtg aaa tgg att cta gga tgt tat gat      354
Cys Pro Gly Ser Thr Gln Leu Val Lys Trp Ile Leu Gly Cys Tyr Asp
        35                  40                  45
gct tta cag aaa aaa tat cta agg atg gtt gtt cta gct gta tac aca      402
Ala Leu Gln Lys Lys Tyr Leu Arg Met Val Val Leu Ala Val Tyr Thr
```

```
    50                  55                  60
aac cca gaa gat cct cag aca att tca gaa tgt tac caa ttc aaa ttc      450
Asn Pro Glu Asp Pro Gln Thr Ile Ser Glu Cys Tyr Gln Phe Lys Phe
65                  70                  75                  80
aaa tac acc aat aat gga cca ctc atg gac ttc ata agt aaa aac caa      498
Lys Tyr Thr Asn Asn Gly Pro Leu Met Asp Phe Ile Ser Lys Asn Gln
                85                  90                  95
agc aac gaa tct agc atg ttg tct act gac acc aag aaa gca agc att      546
Ser Asn Glu Ser Ser Met Leu Ser Thr Asp Thr Lys Lys Ala Ser Ile
            100                 105                 110
ctc ctc att cgc aag att tat atc cta atg caa aat ctg ggg cct tta      594
Leu Leu Ile Arg Lys Ile Tyr Ile Leu Met Gln Asn Leu Gly Pro Leu
        115                 120                 125
cct aat gat gtt tgt ttg acc atg aaa ctt ttt tac tat gat gaa gtt      642
Pro Asn Asp Val Cys Leu Thr Met Lys Leu Phe Tyr Tyr Asp Glu Val
    130                 135                 140
aca ccc cca gat tac cag cct ccc ggt ttt aag gat ggt gat tgt gaa      690
Thr Pro Pro Asp Tyr Gln Pro Pro Gly Phe Lys Asp Gly Asp Cys Glu
145                 150                 155                 160
gga gtt ata ttt gaa ggg gaa cct atg tat tta aat gtg gga gaa gtc      738
Gly Val Ile Phe Glu Gly Glu Pro Met Tyr Leu Asn Val Gly Glu Val
                165                 170                 175
tca aca cct ttt cac atc ttc aaa gta aaa gtg acc act gag aga gaa      786
Ser Thr Pro Phe His Ile Phe Lys Val Lys Val Thr Thr Glu Arg Glu
            180                 185                 190
cga atg gaa aat att gac tca act ata cta tca cca aaa caa ata aaa      834
Arg Met Glu Asn Ile Asp Ser Thr Ile Leu Ser Pro Lys Gln Ile Lys
        195                 200                 205
aca cca ttt caa aaa atc ctg agg gac aaa gat gta gaa gat gaa cag      882
Thr Pro Phe Gln Lys Ile Leu Arg Asp Lys Asp Val Glu Asp Glu Gln
    210                 215                 220
gag cat tat aca agt gat gat ttg gac att gaa act aaa atg gaa gaa      930
Glu His Tyr Thr Ser Asp Asp Leu Asp Ile Glu Thr Lys Met Glu Glu
225                 230                 235                 240
cag gaa aaa aac cct gca tct tct gaa ctt gaa gaa cca agt tta gtt      978
Gln Glu Lys Asn Pro Ala Ser Ser Glu Leu Glu Glu Pro Ser Leu Val
                245                 250                 255
tgt gag gaa gat gaa att atg agg tct aaa gaa agt cca gat ctt tct     1026
Cys Glu Glu Asp Glu Ile Met Arg Ser Lys Glu Ser Pro Asp Leu Ser
            260                 265                 270
att tct cat tct cag gtt gag cag tta gtc aat aaa aca tct gaa ctt     1074
Ile Ser His Ser Gln Val Glu Gln Leu Val Asn Lys Thr Ser Glu Leu
        275                 280                 285
gat atg tct gaa agc aaa aca aga agt gga aaa gtc ttt cag aat aaa     1122
Asp Met Ser Glu Ser Lys Thr Arg Ser Gly Lys Val Phe Gln Asn Lys
    290                 295                 300
atg gca aat gga aat caa cca gta aaa tct tcc aaa gaa aat cgg aag     1170
Met Ala Asn Gly Asn Gln Pro Val Lys Ser Ser Lys Glu Asn Arg Lys
305                 310                 315                 320
aga agt caa cat gaa tct ggg aga ata gtc ctc cat cac ttt gat tct     1218
Arg Ser Gln His Glu Ser Gly Arg Ile Val Leu His His Phe Asp Ser
                325                 330                 335
tct agt caa gag tca gtg cca aaa agg aga aag ttt agt gaa cca aag     1266
Ser Ser Gln Glu Ser Val Pro Lys Arg Arg Lys Phe Ser Glu Pro Lys
            340                 345                 350
gaa cat ata taaaaattat ttttgttctg caggcttgca gagttcttct             1315
Glu His Ile
        355
caccatttaa actgaaggac cctatattat atttccctaa ctctgaagat gtatatgtag   1375
tttaaagcag tttatacact aaaactaagt ttttggctga ctgtcatatt gtggtcctta   1435
atcttgagat aaatccaata gaacttttga ataaaagcaa aagtacaaat gtcataattg   1495
attcggtaat aagtaaaatt tcaaaattga ttttgttcat tacctactta atatttcctt   1555
taaatatata ctaactgtta aggccctcta atgccatttt tctaaacagt aatgtttact   1615
```

```
ttggtattaa aatttggtat tgattcactt tttacttatg ttaaaattat accatttaac   1675
tggctctttt gtcattgtgc tgttattaaa acaatgttct tcaatatttt gacataatgt   1735
attaacattt taatatataa tgtacaattt aagaattggt gctttacctt tactatgctt   1795
tttttacagg acaaaaagac tgatttttaa agtatggcat tttttgcagc ataaataaaa   1855
tattgttcag tacgaaaaaa aaaaaaaaa                                     1884

<210> 19
<211> 691
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 42..515

<220>
<221> sig_peptide
<222> 42..92
<223> Von Heijne matrix
      score 10.7019149919754
      seq VLMLLAVLIWTGA/EN

<400> 19
gagttgtcct gtgctggagg tctgctcaga cgaaggtctc c atg gcg tta gaa gtc     56
                                            Met Ala Leu Glu Val
                                                           -15
ttg atg ctc ctc gct gtc ttg att tgg acc ggt gct gag aac ctc cat      104
Leu Met Leu Leu Ala Val Leu Ile Trp Thr Gly Ala Glu Asn Leu His
        -10                 -5                  1
gtg aaa ata agt tgc tct ctg gac tgg ttg atg gtc tca gtt atc cca      152
Val Lys Ile Ser Cys Ser Leu Asp Trp Leu Met Val Ser Val Ile Pro
5                   10                  15                  20
gtt gca gaa agc aga aat ctg tat ata ttt gcg gat gaa tta cat ctg      200
Val Ala Glu Ser Arg Asn Leu Tyr Ile Phe Ala Asp Glu Leu His Leu
                25                  30                  35
gga atg ggc tgc cct gca aat cgg ata cat aca tat gta tat gag ttt      248
Gly Met Gly Cys Pro Ala Asn Arg Ile His Thr Tyr Val Tyr Glu Phe
            40                  45                  50
ata tat ctt gtt cgt gat tgt ggc atc agg aca agg gta gtt tct gag      296
Ile Tyr Leu Val Arg Asp Cys Gly Ile Arg Thr Arg Val Val Ser Glu
        55                  60                  65
gaa act ctc ctt ttt caa acc gag ctg tac ttt acc cca agg aat ata      344
Glu Thr Leu Leu Phe Gln Thr Glu Leu Tyr Phe Thr Pro Arg Asn Ile
    70                  75                  80
gat cat gac cct cag gaa atc cat ttg gag tgt tcc acc tct agg aaa      392
Asp His Asp Pro Gln Glu Ile His Leu Glu Cys Ser Thr Ser Arg Lys
85                  90                  95                  100
tca gtg tgg ctt aca cca gtt tct act gag aat gaa ata aaa ttg gat      440
Ser Val Trp Leu Thr Pro Val Ser Thr Glu Asn Glu Ile Lys Leu Asp
                105                 110                 115
cct agt cct ttt att gct gac ttt cag aca aca gca gaa gag tta gga      488
Pro Ser Pro Phe Ile Ala Asp Phe Gln Thr Thr Ala Glu Glu Leu Gly
            120                 125                 130
tta tta tct tct agt cca aac ttg ctc tgagctaaag gagaaatgga           535
Leu Leu Ser Ser Ser Pro Asn Leu Leu
        135                 140
aacttgaagc tggtgttatg tattttgcag gaaaacagtt tcatttttc atagcaaaaa    595
tatagttggt gtatatctct ccttaagtct ctggtttcta aaaaccctac ttcagtaaag    655
gtcctgatta gttgattagc gaaaaaaaaa aaaaaa                              691

<210> 20
<211> 1138
<212> DNA
```

<213> Homo sapiens

<220>
<221> CDS
<222> 271..969

<220>
<221> sig_peptide
<222> 271..366
<223> Von Heijne matrix
score 5.6680378526706
seq WMGLACFRSLAAS/SP

<220>
<221> misc_feature
<222> 989
<223> n=a, g, c or t

<400> 20

```
aaaaaccttt caagtgcccc ctccttcct taaagtcttt tatagggtc ccttcttgg      60
ccatctccat cctgtgagtc aggactgaaa gggcacagac aggtcactgc cagcattgtt   120
ggggcaagcc tgcaagcacg catcactggg gatctgacat gacaatggcc gcctgccccc   180
tctgagggct acaggactta ccccagtggg aagcagctaa gcaggtctga ccagccgacc   240
tggacctggc caagggtcct gtcatccctc atg gcc acc ccg cca ttc cgg ctg     294
                                  Met Ala Thr Pro Pro Phe Arg Leu
                                                 -30                 -25
ata agg aag atg ttt tcc ttc aag gtg agc aga tgg atg ggg ctt gcc      342
Ile Arg Lys Met Phe Ser Phe Lys Val Ser Arg Trp Met Gly Leu Ala
                -20                 -15                 -10
tgc ttc cgg tcc ctg gcg gca tcc tct ccc agt att cgc cag aag aaa      390
Cys Phe Arg Ser Leu Ala Ala Ser Ser Pro Ser Ile Arg Gln Lys Lys
            -5                  1                   5
cta atg cac aag ctg cag gag gaa aag gct ttt cgc gaa gag atg aaa      438
Leu Met His Lys Leu Gln Glu Glu Lys Ala Phe Arg Glu Glu Met Lys
    10                  15                  20
att ttt cgt gaa aaa ata gag gac ttc agg gaa gag atg tgg act ttc      486
Ile Phe Arg Glu Lys Ile Glu Asp Phe Arg Glu Glu Met Trp Thr Phe
25                  30                  35                  40
cga ggc aag atc cat gct ttc cgg ggc cag atc ctg ggt ttt tgg gaa      534
Arg Gly Lys Ile His Ala Phe Arg Gly Gln Ile Leu Gly Phe Trp Glu
                45                  50                  55
gag gag aga cct ttc tgg gaa gag gag aaa acc ttc tgg aaa gag gaa      582
Glu Glu Arg Pro Phe Trp Glu Glu Glu Lys Thr Phe Trp Lys Glu Glu
            60                  65                  70
aaa tcc ttc tgg gaa atg gaa aag tct ttc agg gag gaa gag aaa act      630
Lys Ser Phe Trp Glu Met Glu Lys Ser Phe Arg Glu Glu Glu Lys Thr
        75                  80                  85
ttc tgg aaa aag tac cgc act ttc tgg aag gag gat aag gcc ttc tgg      678
Phe Trp Lys Lys Tyr Arg Thr Phe Trp Lys Glu Asp Lys Ala Phe Trp
    90                  95                  100
aaa gag gac aat gcc tta tgg gaa aga gac cgg aac ctt ctt cag gag      726
Lys Glu Asp Asn Ala Leu Trp Glu Arg Asp Arg Asn Leu Leu Gln Glu
105                 110                 115                 120
gac aag gcc ctg tgg gag gaa gaa aag gcc ctg tgg gta gag gaa aga      774
Asp Lys Ala Leu Trp Glu Glu Glu Lys Ala Leu Trp Val Glu Glu Arg
                125                 130                 135
gcc ctc ctt gag ggg gag aaa gcc ctg tgg gaa gat aaa acg tcc ctc      822
Ala Leu Leu Glu Gly Glu Lys Ala Leu Trp Glu Asp Lys Thr Ser Leu
            140                 145                 150
tgg gag gaa gag aat gcc ctc tgg gag gaa gag agg gcc ttc tgg atg      870
Trp Glu Glu Glu Asn Ala Leu Trp Glu Glu Glu Arg Ala Phe Trp Met
        155                 160                 165
gag aac aat ggc cac att gcc gga gag cag atg ctc gaa gat ggg ccc      918
```

```
Glu Asn Asn Gly His Ile Ala Gly Glu Gln Met Leu Glu Asp Gly Pro
    170                 175                 180
cac aac gcc aac aga ggg cag cgc ttg ctg gcc ttc tcc cga ggc agg      966
His Asn Ala Asn Arg Gly Gln Arg Leu Leu Ala Phe Ser Arg Gly Arg
185                 190                 195                 200
gcg tagccagcat gcaggtgcan gggccctgtg gtccagactc ccctgggttg          1019
Ala
ggattcaagt ccagggtgag cccatgtgct ggagaaaata cacactcatt ggtctccttg   1079
ctttgaaaga tccaataaag tcctgaggca aggtttggaa aaccaaaaaa aaaaaaaaa    1138

<210> 21
<211> 468
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 76..276

<220>
<221> sig_peptide
<222> 76..135
<223> Von Heijne matrix
      score 5.21332530399231
      seq SPVFLVFPPEITA/SE

<400> 21
agcacaagaa aagaacatgg tctagactga agtaccaact aaatcatctc ctttcaaatt     60
atcaccgaca ccatc atg gat tca agc acc gca cac agt ccg gtg ttt ctg     111
                Met Asp Ser Ser Thr Ala His Ser Pro Val Phe Leu
                -20                 -15                 -10
gta ttt cct cca gaa atc act gct tca gaa tat gag tcc aca gaa ctt      159
Val Phe Pro Pro Glu Ile Thr Ala Ser Glu Tyr Glu Ser Thr Glu Leu
            -5                  1               5
tca gcc acg acc ttt tca act caa agc ccc ttg caa aaa tta ttt gct      207
Ser Ala Thr Thr Phe Ser Thr Gln Ser Pro Leu Gln Lys Leu Phe Ala
    10                  15                  20
aga aaa atg aaa atc tta ggg gat atc cat tct ggg gct ctg ttt tgt      255
Arg Lys Met Lys Ile Leu Gly Asp Ile His Ser Gly Ala Leu Phe Cys
25                  30                  35                  40
tca tta att ctg gag cct tcc taattgcagt gaaaagaaaa accacagaaa         306
Ser Leu Ile Leu Glu Pro Ser
                45
ctctgggaat tttgattaca ttgatgactt tcagcattat tgaattattc atttctctgc    366
ctttctcaat tttggggtgc cactcagagg attgtgattg tgaacaatgt tgttgactag    426
cactgtgaga ataaagatgt gttaaaataa aaaaaaaaaa aa                       468

<210> 22
<211> 720
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 6..287

<220>
<221> sig_peptide
<222> 6..80
<223> Von Heijne matrix
      score 4.17710408129886
      seq ISLSHLFLDLSRS/LW
```

<400> 22
atttg atg tgc ttc tta gtc tcg ttt aac ttg ccg att cat ata tcc ctg 50
Met Cys Phe Leu Val Ser Phe Asn Leu Pro Ile His Ile Ser Leu
-25 -20 -15
tct cat ttg ttc tta gat ttg tca cga agc ctc tgg ttt ttg gct tgt 98
Ser His Leu Phe Leu Asp Leu Ser Arg Ser Leu Trp Phe Leu Ala Cys
-10 -5 1 5
cct ggt ttg aac ttg gtg tat ctg gct ctt gac tca ttt tct gac ctc 146
Pro Gly Leu Asn Leu Val Tyr Leu Ala Leu Asp Ser Phe Ser Asp Leu
10 15 20
aga cca tcc tta aat ctg ctt ttc tac ttt gta cca ggc ttt ggc gtc 194
Arg Pro Ser Leu Asn Leu Leu Phe Tyr Phe Val Pro Gly Phe Gly Val
25 30 35
tcc aag tac ctg acc tca gct caa cct gtc ttg ggt ttt ctt ctc ctc 242
Ser Lys Tyr Leu Thr Ser Ala Gln Pro Val Leu Gly Phe Leu Leu Leu
40 45 50
cct gac att gac aac cca gcc ctc cta ggc aca gag aga tgg agc 287
Pro Asp Ile Asp Asn Pro Ala Leu Leu Gly Thr Glu Arg Trp Ser
55 60 65
tgagtgtggt tttcctgaaa taaagcttgc attatgagag ggaataaaca gaagaaaaaa 347
atagtaagta aaatcttgct tgcctctcag taaaataaag ctctattttt cgttttttt 407
ttttccaact tcctgtacaa aaaagggaaa actttagctt ttgggggaaa tttggagcta 467
gcctgttggt actgttgagc ttagtgtatc tataactata tattattcca caatatctta 527
aatactttat aaagatattt tcataaatta cagcaatcct ggctttagat gattgatggc 587
catttttaaa caattaaagc taatttctag ctttttatga gtttggtatt aagcacagta 647
gtttcttaga aagtctccag ggaatgcatt ttgcaaaata aaaatcagct aatgacccaa 707
aaaaaaaaaa aaa 720

<210> 23
<211> 727
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 171..692

<220>
<221> sig_peptide
<222> 171..227
<223> Von Heijne matrix
score 4.17573075349936
seq LLLGQRCSLKVSG/QE

<400> 23
attgtgacat caccgtgcac tagccaatgg ctgcctgcct aagctgggtc cctggtctcc 60
tgggactact agccctttgt tgatagggag aagccaacat ctcccgcagg acccctaat 120
cttcagggca gctcccagag catggatccc tcctgattcc actcagcccg atg ttc 176
Met Phe
ctc aca gtc aag ctg ctc ctg ggc cag aga tgc agt ctg aag gtg tca 224
Leu Thr Val Lys Leu Leu Leu Gly Gln Arg Cys Ser Leu Lys Val Ser
-15 -10 -5
ggg caa gag agt gta gcc acg ctg aag aga ctg gtg tcc agg cgg ctg 272
Gly Gln Glu Ser Val Ala Thr Leu Lys Arg Leu Val Ser Arg Arg Leu
1 5 10 15
aag gtg cct gag gag cag cag cac ctg ctt ttc cgt ggc cag ctc ctg 320
Lys Val Pro Glu Glu Gln Gln His Leu Leu Phe Arg Gly Gln Leu Leu
20 25 30
gag gat gac aag cac ctc tct gac tac tgc att ggg ccc aat gcc tct 368
Glu Asp Asp Lys His Leu Ser Asp Tyr Cys Ile Gly Pro Asn Ala Ser
35 40 45
atc aat gtc atc atg cag ccc ttg gag aag atg gcg cta aag gag gcc 416
Ile Asn Val Ile Met Gln Pro Leu Glu Lys Met Ala Leu Lys Glu Ala

```
          50                  55                  60
   cac cag ccg cag acc cag ccc ctg tgg cac cag ctg gga ctg gtc cta      464
   His Gln Pro Gln Thr Gln Pro Leu Trp His Gln Leu Gly Leu Val Leu
       65                  70                  75
   gct aaa cac ttt gaa cca cag gat gcc aag gcc gtg ctg cag ctg cta      512
   Ala Lys His Phe Glu Pro Gln Asp Ala Lys Ala Val Leu Gln Leu Leu
   80                  85                  90                  95
   agg cag gag cac gag gag cgc ctg cag aag ata agc ctg gag cac ctg      560
   Arg Gln Glu His Glu Glu Arg Leu Gln Lys Ile Ser Leu Glu His Leu
                   100                 105                 110
   gag cag ctg gcc cag tac ctc ctg gca gag gag cct cac gtg gag cca      608
   Glu Gln Leu Ala Gln Tyr Leu Leu Ala Glu Glu Pro His Val Glu Pro
               115                 120                 125
   gct gga gag agg gag ctt gag gcg aag gca cgg cct cag agc tcc tgt      656
   Ala Gly Glu Arg Glu Leu Glu Ala Lys Ala Arg Pro Gln Ser Ser Cys
           130                 135                 140
   gac atg gag gag aag gag gag gca gca gct gat cag taaacgggcc           702
   Asp Met Glu Glu Lys Glu Glu Ala Ala Ala Asp Gln
       145                 150                 155
   atcctacccg aaaaaaaaaa aaaaa                                        727

   <210> 24
   <211> 470
   <212> DNA
   <213> Homo sapiens

   <220>
   <221> CDS
   <222> 137..454

   <220>
   <221> sig_peptide
   <222> 137..187
   <223> Von Heijne matrix
         score 10.7019149919754
         seq VLMLLAVLIWTGA/EN

   <400> 24
   atcctgtgaa ctacccaaaa ggaggaaaac gaacgcagct gagcatggga tgccatataa     60
   aaatcactta aaccagtcgc cactccttgt ttcctgagtt gtcctgtgct ggaggtctgc    120
   tcagacgaag gtctcc atg gcg tta gaa gtc ttg atg ctc ctc gct gtc ttg    172
                     Met Ala Leu Glu Val Leu Met Leu Leu Ala Val Leu
                            -15                 -10
   att tgg acc ggt gct gag aac ctc cat gtg aaa ata agt tgc tct ctg      220
   Ile Trp Thr Gly Ala Glu Asn Leu His Val Lys Ile Ser Cys Ser Leu
   -5                  1                   5                  10
   gac tgg ttg atg gtc tca gtt atc cca gtt gca gaa agc aga aat ctg      268
   Asp Trp Leu Met Val Ser Val Ile Pro Val Ala Glu Ser Arg Asn Leu
               15                  20                  25
   tat ata ttt gcg gat gaa tta cat ctg gga atg ggc tgc cct gca aat      316
   Tyr Ile Phe Ala Asp Glu Leu His Leu Gly Met Gly Cys Pro Ala Asn
           30                  35                  40
   cgg ata cat aca tat gta tat gag ttt ata tat ctt gtt cgt gat tgt      364
   Arg Ile His Thr Tyr Val Tyr Glu Phe Ile Tyr Leu Val Arg Asp Cys
       45                  50                  55
   ggc atc agg aca agg gta aga aca gtg att gtc tgt aaa aaa tac tgc      412
   Gly Ile Arg Thr Arg Val Arg Thr Val Ile Val Cys Lys Lys Tyr Cys
   60                  65                  70                  75
   atg ttt tgt cag act ttt atg cct agt att aaa att gtc ttt              454
   Met Phe Cys Gln Thr Phe Met Pro Ser Ile Lys Ile Val Phe
                   80                  85
   taaaaaaaaa aaaaaa                                                  470
```

<210> 25
<211> 987
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 238..609

<220>
<221> sig_peptide
<222> 238..291
<223> Von Heijne matrix
score 10.0374888212272
seq LLLLVMALPPGTT/GV

<400> 25
attccattca cagactcttg ttgggcagca gccacccgct cacctccatc cccaggactt 60
agagggacgc agggcgttgg gaacagagga cactccaggc gctgaccctg ggaggccagg 120
accagggcca aagtcccgtg ggcaagagga gtcctcagag gtccttcatt cagcggttcc 180
gggaggtctg ggaagcccac ggcctggctg gggcagggtc aacgccgcca ggccgcc 237
atg gtc ctg tgc tgg ctg ctg ctt ctg gtg atg gct ctg ccc cca ggc 285
Met Val Leu Cys Trp Leu Leu Leu Leu Val Met Ala Leu Pro Pro Gly
-15 -10 -5
acg acg ggc gtc aag gac tgc gtc ttc tgt gag ctc acc gac tcc atg 333
Thr Thr Gly Val Lys Asp Cys Val Phe Cys Glu Leu Thr Asp Ser Met
1 5 10
cag tgt cct ggt acc tac atg cac tgt ggc gat gac gag gac tgc ttc 381
Gln Cys Pro Gly Thr Tyr Met His Cys Gly Asp Asp Glu Asp Cys Phe
15 20 25 30
aca ggc cac ggg gtc gcc ccg ggc act ggt ccg gtc atc aac aaa ggc 429
Thr Gly His Gly Val Ala Pro Gly Thr Gly Pro Val Ile Asn Lys Gly
35 40 45
tgc ctg cga gcc acc agc tgc ggc ctt gag gaa ccc gtc agc tac agg 477
Cys Leu Arg Ala Thr Ser Cys Gly Leu Glu Glu Pro Val Ser Tyr Arg
50 55 60
ggc gtc acc tac agc ctc acc acc aac tgc tgc acc ggc cgc ctg tgt 525
Gly Val Thr Tyr Ser Leu Thr Thr Asn Cys Cys Thr Gly Arg Leu Cys
65 70 75
aac aga gcc ccg agc agc cag aca gtg ggg gcc acc acc agc ctg gca 573
Asn Arg Ala Pro Ser Ser Gln Thr Val Gly Ala Thr Thr Ser Leu Ala
80 85 90
ctg ggg ctg ggt atg ctg ctt cct cca cgt ttg ctg tgaccaacag 619
Leu Gly Leu Gly Met Leu Leu Pro Pro Arg Leu Leu
95 100 105
ggaggacagg gcctgggact gttctcccag atccgccact ccccatgtcc ccatgtcctt 679
cccccactaa atggccagag aggccctgga caacctcttg cggccctggc ttcatccctt 739
ctaaggctgt ccaccaggag cccggtgcta ggggaagcat ccccaggcct gactgagcgg 799
caggggagca cggcccgtgg gtttgattgt attactctgt tccactggtt ctaagacgca 859
gagcttctca catctcaatc aggatgcttc tctccattgg tagcacttta gagtccatga 919
aatatggtaa aaaatatata tatatcataa taaatgacag ctgatgttca tggaaaaaaa 979
aaaaaaaa 987

<210> 26
<211> 908
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 80..862

<220>

<221> sig_peptide
<222> 80..127
<223> Von Heijne matrix
score 3.66725851505537
seq FSLLSISGPPISS/SA

<400> 26
gaatgtttat cctctggaca aaccagccag cctctccaga gcaggcgtgt gatctctgta 60
cccccgcagt ggtcagaat atg gag aac ttc tca ctc ctc agc atc tct gga 112
Met Glu Asn Phe Ser Leu Leu Ser Ile Ser Gly
-15 -10
cct cca atc tct tcc tcc gcc ctg agt gct ttt ccc gac att atg ttc 160
Pro Pro Ile Ser Ser Ser Ala Leu Ser Ala Phe Pro Asp Ile Met Phe
-5 1 5 10
tct cgt gcc acc agc ctg cca gac att gca aag aca gca gta ccc act 208
Ser Arg Ala Thr Ser Leu Pro Asp Ile Ala Lys Thr Ala Val Pro Thr
15 20 25
gag gca tcc agc cca gct cag gcc ctg cca ccc cag tac caa agc atc 256
Glu Ala Ser Ser Pro Ala Gln Ala Leu Pro Pro Gln Tyr Gln Ser Ile
30 35 40
att gtc agg caa ggg ata cag aac aca gtg ctc tca cca gac tgc agc 304
Ile Val Arg Gln Gly Ile Gln Asn Thr Val Leu Ser Pro Asp Cys Ser
45 50 55
ttg ggg gac acc cag cac gga gag aag ctg agg cgg aac tgc act atc 352
Leu Gly Asp Thr Gln His Gly Glu Lys Leu Arg Arg Asn Cys Thr Ile
60 65 70 75
tac cgg ccc tgg ttc tcc ccc tac agc tac ttc gtg tgt gca gac aaa 400
Tyr Arg Pro Trp Phe Ser Pro Tyr Ser Tyr Phe Val Cys Ala Asp Lys
80 85 90
gag agc cag ctg gag gcc tat gac ttc cca gag gtg cag cag gat gag 448
Glu Ser Gln Leu Glu Ala Tyr Asp Phe Pro Glu Val Gln Gln Asp Glu
95 100 105
ggc aag tgg gac aac tgc ctt tct gag gac atg gct gag aac atc tgt 496
Gly Lys Trp Asp Asn Cys Leu Ser Glu Asp Met Ala Glu Asn Ile Cys
110 115 120
tcg tcc tct tcc tcc cca gag aac act tgc cct cga gaa gcc acc aag 544
Ser Ser Ser Ser Ser Pro Glu Asn Thr Cys Pro Arg Glu Ala Thr Lys
125 130 135
aaa tcc agg cat ggc ctg gac tcc atc aca tcc cag gac atc cta atg 592
Lys Ser Arg His Gly Leu Asp Ser Ile Thr Ser Gln Asp Ile Leu Met
140 145 150 155
gct tcc aga tgg cac cca gca cag cag aat ggc tac aag tgc gtg gcc 640
Ala Ser Arg Trp His Pro Ala Gln Gln Asn Gly Tyr Lys Cys Val Ala
160 165 170
tgc tgc cgc atg tac ccc acc ctg gac ttc ctc aag agc cac atc aag 688
Cys Cys Arg Met Tyr Pro Thr Leu Asp Phe Leu Lys Ser His Ile Lys
175 180 185
agg ggc ttc agg gag ggc ttc agc tgc aag gtg tac tac cgc aag ctc 736
Arg Gly Phe Arg Glu Gly Phe Ser Cys Lys Val Tyr Tyr Arg Lys Leu
190 195 200
aaa gcc ctc tgg agc aag gag cag aag gcc cgg ctg gga gac agg ctc 784
Lys Ala Leu Trp Ser Lys Glu Gln Lys Ala Arg Leu Gly Asp Arg Leu
205 210 215
tcc tcc ggc agc tgc cag gcc ttc aat agt cct gct gaa cac ctt agg 832
Ser Ser Gly Ser Cys Gln Ala Phe Asn Ser Pro Ala Glu His Leu Arg
220 225 230 235
caa att ggc ggt gaa gcc tac tta tgt ctc tagagagatg ccaataaagt 882
Gln Ile Gly Gly Glu Ala Tyr Leu Cys Leu
240 245
tagtcacagc caaaaaaaaa aaaaaa 908

<210> 27
<211> 762

<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 83..310

<220>
<221> sig_peptide
<222> 83..157
<223> Von Heijne matrix
score 4.72955689475746
seq LCALLSNFCPSTT/VK

<400> 27
ttttttctac tacaaacgcc atggggatgc gggtctggga acagcggaaa accctaccct 60
gccctgaaaa gtccctggct ca atg tgc atg tcc ctt tct atg aaa gtt cct 112
Met Cys Met Ser Leu Ser Met Lys Val Pro
-25 -20
tgc tgc cta tgc gcc ttg ctc tct aac ttc tgt ccc tcc aca act gtg 160
Cys Cys Leu Cys Ala Leu Leu Ser Asn Phe Cys Pro Ser Thr Thr Val
-15 -10 -5 1
aaa gga gac gtc gtg act tcc ttc ttt cgt gct gac tat gac tta gcc 208
Lys Gly Asp Val Val Thr Ser Phe Phe Arg Ala Asp Tyr Asp Leu Ala
5 10 15
agt agg tct gca gat cag tcc tcc cag aaa gtg aag ttg cgc atg ttc 256
Ser Arg Ser Ala Asp Gln Ser Ser Gln Lys Val Lys Leu Arg Met Phe
20 25 30
act ggg cgt ctt ccc atc ggc ccc ttc gcc agt gtg ggg aac gcg gcg 304
Thr Gly Arg Leu Pro Ile Gly Pro Phe Ala Ser Val Gly Asn Ala Ala
35 40 45
gag ctg tgagccggcg actcgggtcc ctgaggtctg gattctttct ccgctactga 360
Glu Leu
50
gacacggcgg acacacacaa acacagaacc acacagccag tcccaggagc ccagtaatgg 420
agagccccaa aaagaagaac cagcagctga aagtcgggat cctacacctg ggcagcagac 480
agaagaagat caggatacag ctgagatccc agtgcgcgac atggaaggtg atctgcaaga 540
gctgcatcag tcaaacaccg gggataaatc tggatttggg ttccggcgtc aaggtgaaga 600
taatacctaa agaggaacac tgtaaaatgc cagaagcagg tgaagagcaa ccacaagttt 660
aaatgaagac aagctgaaac aacgcaagct ggttttatat tagatatttg acttaaacta 720
tctcaataaa gttttgcagc tttcaccaaa aaaaaaaaaa aa 762

<210> 28
<211> 1102
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 310..906

<220>
<221> sig_peptide
<222> 310..357
<223> Von Heijne matrix
score 11.0931109030915
seq FPLLLLSLGLVLA/EA

<400> 28
atacagtgac ctagagcagg catgggtggg tcacaggctt tggagagcac tctctgtcct 60
gatcttttca gttgagagac ttcagctgtt cattgctcat ttggacttag ttcaaggtca 120
tgtcaaagaa gaaggtgcac ttacgctagt tgttagctct gtcttttgta accatcaagt 180
tccatgcgat tgatcagatt taggaggggg cgttggggga taatcaattt tgggtgtcac 240

```
caggtaaaca gagccctcag catctgaata gaaactgaac aggaacagaa gagattcact    300
acatctgag atg gag acc ttt cct ctg ctg ctg ctc agc ctg ggc ctg gtt    351
          Met Glu Thr Phe Pro Leu Leu Leu Leu Ser Leu Gly Leu Val
              -15                 -10                 -5
ctt gca gaa gca tca gaa agc aca atg aag ata att aaa gaa gaa ttt      399
Leu Ala Glu Ala Ser Glu Ser Thr Met Lys Ile Ile Lys Glu Glu Phe
        1               5                  10
aca gac gaa gag atg caa tat gac atg gca aaa agt ggc caa gaa aaa      447
Thr Asp Glu Glu Met Gln Tyr Asp Met Ala Lys Ser Gly Gln Glu Lys
15                  20                  25                  30
cag acc att gag ata tta atg aac ccg atc ctg tta gtt aaa aat acc      495
Gln Thr Ile Glu Ile Leu Met Asn Pro Ile Leu Leu Val Lys Asn Thr
                35                  40                  45
agc ctc agc atg tcc aag gat gat atg tct tcc aca tta ctg aca ttc      543
Ser Leu Ser Met Ser Lys Asp Asp Met Ser Ser Thr Leu Leu Thr Phe
            50                  55                  60
aga agt tta cat tat aat gac ccc aag gga aac agt tcg ggt aat gac      591
Arg Ser Leu His Tyr Asn Asp Pro Lys Gly Asn Ser Ser Gly Asn Asp
        65                  70                  75
aaa gag tgt tgc aat gac atg aca gtc tgg aga aaa gtt tca gaa gca      639
Lys Glu Cys Cys Asn Asp Met Thr Val Trp Arg Lys Val Ser Glu Ala
    80                  85                  90
aac gga tcg tgc aag tgg agc aat aac ttc atc cgc agc tcc aca gaa      687
Asn Gly Ser Cys Lys Trp Ser Asn Asn Phe Ile Arg Ser Ser Thr Glu
95                  100                 105                 110
gtg atg cgc agg gtc cac agg gcc ccc agc tgc aag ttt gta cag aat      735
Val Met Arg Arg Val His Arg Ala Pro Ser Cys Lys Phe Val Gln Asn
                115                 120                 125
cct ggc ata agc tgc tgt gag agc cta gaa ctg gaa aat aca gtg tgc      783
Pro Gly Ile Ser Cys Cys Glu Ser Leu Glu Leu Glu Asn Thr Val Cys
            130                 135                 140
cag ttc act aca ggc aaa caa ttc ccc agg tgc caa tac cat agt gtt      831
Gln Phe Thr Thr Gly Lys Gln Phe Pro Arg Cys Gln Tyr His Ser Val
        145                 150                 155
acc tca tta gag aag ata ttg aca gtg ctg aca ggt cat tct ctg atg      879
Thr Ser Leu Glu Lys Ile Leu Thr Val Leu Thr Gly His Ser Leu Met
    160                 165                 170
agc tgg tta gtt tgt ggc tct aag ttg taaatcccac agagctttag            926
Ser Trp Leu Val Cys Gly Ser Lys Leu
175                 180
gactagggtc ttactaaaga aggacctctt cttgttcatt cttgtttaaa cctttcctta    986
atatctactc tttagcacta tagtgaactc ctgattattt attctaactg gaggagtgaa   1046
aaatccaaaa ttgtggataa ttcaattaaa agttatgact gaaaaaaaaa aaaaaa       1102

<210> 29
<211> 436
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 24..287

<220>
<221> sig_peptide
<222> 24..131
<223> Von Heijne matrix
      score 3.79790641648006
      seq ILMRDFSPSGIFG/AF

<400> 29
acagcggaca ccaggactcc aaa atg gcg tca gtt gta cca gtg aag gac aag     53
                          Met Ala Ser Val Val Pro Val Lys Asp Lys
```

```
                                    -35                 -30
aaa ctt ctg gag gtc aaa ctg ggg gag ctg cca agc tgg atc ttg atg      101
Lys Leu Leu Glu Val Lys Leu Gly Glu Leu Pro Ser Trp Ile Leu Met
    -25                 -20                 -15
cgg gac ttc agt cct agt ggc att ttc gga gcg ttt caa aga ggt tac      149
Arg Asp Phe Ser Pro Ser Gly Ile Phe Gly Ala Phe Gln Arg Gly Tyr
-10                 -5                  1                   5
tac cgg tac tac aac aag tac atc aat gtg aag aag ggg agc atc tcg      197
Tyr Arg Tyr Tyr Asn Lys Tyr Ile Asn Val Lys Lys Gly Ser Ile Ser
            10                  15                  20
ggg att acc atg gtg ctg gca tgc tac gtg ctc ttt agc tac tcc ttt      245
Gly Ile Thr Met Val Leu Ala Cys Tyr Val Leu Phe Ser Tyr Ser Phe
        25                  30                  35
tcc tac aag cat ctc aag cac gag cgg ctc cgc aaa tac cac              287
Ser Tyr Lys His Leu Lys His Glu Arg Leu Arg Lys Tyr His
    40                  45                  50
tgaagaggac acactctgca ccccccccacc ccacgacctt ggcccgagcc cctccgtgag    347
gaacacaatc tcaatcgttg ctgaatcctt tcatatccta ataggaatta acctccaaat    407
aaaacatgac tggtaaaaaa aaaaaaaaa                                      436

<210> 30
<211> 1938
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 132..1574

<220>
<221> sig_peptide
<222> 132..206
<223> Von Heijne matrix
      score 11.1130239236827
      seq LALLLTSTPEALG/AN

<400> 30
ctccccttcc cgctcccagg aacccatcca gcctcaggaa ctgcccccag ccatcgagcc     60
ttggctactt aagggacctg ggcccaatcc acagctggga cagtcctggc ccactgcact    120
gggaatctag g atg ggg gcc ttg gcc aga gcc ctg ccg tcc ata ctg ctg      170
             Met Gly Ala Leu Ala Arg Ala Leu Pro Ser Ile Leu Leu
             -25                 -20                 -15
gca ttg ctg ctt acg tcc acc cca gag gct ctg ggt gcc aac ccc ggc      218
Ala Leu Leu Leu Thr Ser Thr Pro Glu Ala Leu Gly Ala Asn Pro Gly
        -10                 -5                  1
ttg gtc gcc agg atc acc gac aag gga ctg cag tat gcg gcc cag gag      266
Leu Val Ala Arg Ile Thr Asp Lys Gly Leu Gln Tyr Ala Ala Gln Glu
5                   10                  15                  20
ggg cta ttg gct ctg cag agt gag ctg ctc agg atc acg ctg cct gac      314
Gly Leu Leu Ala Leu Gln Ser Glu Leu Leu Arg Ile Thr Leu Pro Asp
                25                  30                  35
ttc acc ggg gac ttg agg atc ccc cac gtc ggc cgt ggg cgc tat gag      362
Phe Thr Gly Asp Leu Arg Ile Pro His Val Gly Arg Gly Arg Tyr Glu
            40                  45                  50
ttc cac agc ctg aac atc cac agc tgt gag ctg ctt cac tct gcg ctg      410
Phe His Ser Leu Asn Ile His Ser Cys Glu Leu Leu His Ser Ala Leu
        55                  60                  65
agg cct gtc cct ggc cag ggc ctg agt ctc agc atc tcc gac tcc tcc      458
Arg Pro Val Pro Gly Gln Gly Leu Ser Leu Ser Ile Ser Asp Ser Ser
    70                  75                  80
atc cgg gtc cag ggc agg tgg aag gtg cgc aag tca ttc ttc aaa cta      506
Ile Arg Val Gln Gly Arg Trp Lys Val Arg Lys Ser Phe Phe Lys Leu
85                  90                  95                  100
```

```
cag ggc tcc ttt gat gtc agt gtc aag ggc atc agc att tcg gtc aac      554
Gln Gly Ser Phe Asp Val Ser Val Lys Gly Ile Ser Ile Ser Val Asn
                105                 110                 115
ctc ctg ttg ggc agc gat tcc tcc ggg agg ccc aca gtt act gcc tcc      602
Leu Leu Leu Gly Ser Asp Ser Ser Gly Arg Pro Thr Val Thr Ala Ser
            120                 125                 130
agc tgc agc agt gac atc gct gac gtg gag gtg gac atg tcg gga gac      650
Ser Cys Ser Ser Asp Ile Ala Asp Val Glu Val Asp Met Ser Gly Asp
        135                 140                 145
ttg ggg tgg ctg ttg aac ctc ttc cac aac cag att gag tcc aag ttc      698
Leu Gly Trp Leu Leu Asn Leu Phe His Asn Gln Ile Glu Ser Lys Phe
    150                 155                 160
cag aaa gta ctg gag agc agg att tgc gaa atg atc cag aaa tcg gtg      746
Gln Lys Val Leu Glu Ser Arg Ile Cys Glu Met Ile Gln Lys Ser Val
165                 170                 175                 180
tcc tcc gat cta cag cct tat ctc caa act ctg aca gtt aca aca gag      794
Ser Ser Asp Leu Gln Pro Tyr Leu Gln Thr Leu Thr Val Thr Thr Glu
                185                 190                 195
att gac agt ttc gcc gac att gat tat agc tta gtg gaa gcc cct cgg      842
Ile Asp Ser Phe Ala Asp Ile Asp Tyr Ser Leu Val Glu Ala Pro Arg
            200                 205                 210
gca aca gcc cag atg ctg gag gtg atg ttt aag ggt gaa atc ttt cat      890
Ala Thr Ala Gln Met Leu Glu Val Met Phe Lys Gly Glu Ile Phe His
        215                 220                 225
cgt aac cac cgt tct cca gtt acc ctc ctt gct gca gtc atg agc ctt      938
Arg Asn His Arg Ser Pro Val Thr Leu Leu Ala Ala Val Met Ser Leu
    230                 235                 240
cct gag gaa cac aac aaa atg gtc tac ttt gcc atc tcg gat tat gtc      986
Pro Glu Glu His Asn Lys Met Val Tyr Phe Ala Ile Ser Asp Tyr Val
245                 250                 255                 260
ttc aac acg gcc agc ctg gtt tat cat gag gaa gga tat ctg aac ttc     1034
Phe Asn Thr Ala Ser Leu Val Tyr His Glu Glu Gly Tyr Leu Asn Phe
                265                 270                 275
tcc atc aca gat gac atg ata ccg cct gac tct aat atc cga ctg acc     1082
Ser Ile Thr Asp Asp Met Ile Pro Pro Asp Ser Asn Ile Arg Leu Thr
            280                 285                 290
acc aag tcc ttc cga ccc ttc gtc cca cgg tta gcc agg ctc tac ccc     1130
Thr Lys Ser Phe Arg Pro Phe Val Pro Arg Leu Ala Arg Leu Tyr Pro
        295                 300                 305
aac atg aac ctg gaa ctc cag gga tca gtg ccc tct gct ccg ctc ctg     1178
Asn Met Asn Leu Glu Leu Gln Gly Ser Val Pro Ser Ala Pro Leu Leu
    310                 315                 320
aac ttc agc cct ggg aat ctg tct gtg gac ccc tat atg gag ata gat     1226
Asn Phe Ser Pro Gly Asn Leu Ser Val Asp Pro Tyr Met Glu Ile Asp
325                 330                 335                 340
gcc ttt gtg ctc ctg ccc agc tcc agc aag gag cct gtc ttc cgg ctc     1274
Ala Phe Val Leu Leu Pro Ser Ser Ser Lys Glu Pro Val Phe Arg Leu
                345                 350                 355
agt gtg gcc act aat gtg tcc gcc acc ttg acc ttc aat acc agc aag     1322
Ser Val Ala Thr Asn Val Ser Ala Thr Leu Thr Phe Asn Thr Ser Lys
            360                 365                 370
atc act ggg ttc ctg aag cca gga aag gta aaa gtg gaa ctg aaa gaa     1370
Ile Thr Gly Phe Leu Lys Pro Gly Lys Val Lys Val Glu Leu Lys Glu
        375                 380                 385
tcc aaa gtt gga cta ttc aat gca gag ctg ttg gaa gcg ctc ctc aac     1418
Ser Lys Val Gly Leu Phe Asn Ala Glu Leu Leu Glu Ala Leu Leu Asn
    390                 395                 400
tat tac atc ctt aac acc ttc tac ccc aag ttc aat gat aag ttg gcc     1466
Tyr Tyr Ile Leu Asn Thr Phe Tyr Pro Lys Phe Asn Asp Lys Leu Ala
405                 410                 415                 420
gaa ggc ttc ccc ctt cct ctg ctg aag cgt gtt cag ctc tac gac ctt     1514
Glu Gly Phe Pro Leu Pro Leu Leu Lys Arg Val Gln Leu Tyr Asp Leu
                425                 430                 435
```

```
ggg ctg cag atc cat aag gac ttc ctg ttc ttg ggt gcc aat gtc caa     1562
Gly Leu Gln Ile His Lys Asp Phe Leu Phe Leu Gly Ala Asn Val Gln
            440                 445                 450
tac atg aga gtt tgaggacaag aaagatgaag cttggaggtc acagctggat         1614
Tyr Met Arg Val
        455
ctgcttgttg catttccagc tgtgcagcac gtctcagaga ttcttgaaga atgaagacat   1674
ttctgctctc agctccgggg gtgaggtgtg cctggcctct gcctccaccc tcctcctctt   1734
caccaggtgc atgcatgccc tctctgagtc tggactttgc ttcccctcca ggagggacca   1794
ccctccctga ctggcctggg atatctttac aagcaggcac tgtatttttt tattcgccat   1854
ctgatcccca tgcctagcag agtgctggca cttagtaggt cctcaataaa tatttattaa   1914
atgatgacaa aaaaaaaaaa aaaa                                          1938

<210> 31
<211> 1116
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 117..545

<220>
<221> sig_peptide
<222> 117..245
<223> Von Heijne matrix
      score 5.65876793443964
      seq VVSFALIATLVYA/LF

<400> 31
ataaggggac gtctagtggg ttgcccggga ggggtggcgg gagcggtcct ggaaataatc     60
tgtcctctgt cgccgggaac tggcgaggta gttccttcgc ggtggagaga cctgga atg     119
                                                              Met
gcc aaa tat caa ggt gaa gtt caa agt ttg aaa ctg gat gat gat tca       167
Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp Ser
        -40                 -35                 -30
gtt ata gaa gga gta agc gac caa gta ctt gtg gca gtt gtg gtc agt       215
Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val Ser
    -25                 -20                 -15
ttc gct ttg att gct acc ctg gta tat gca ctt ttc aga aat gta cat       263
Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val His
-10                 -5                  1                   5
caa aac att cac cca gaa aac cag gag cta gta agg gta ctt cga gaa       311
Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg Glu
            10                  15                  20
cag ctt caa aca gaa cag gat gca cct gct gcc act cga cag cag ttc       359
Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Ala Thr Arg Gln Gln Phe
        25                  30                  35
tac act gac atg tac tgt ccc atc tgc ctg cac caa gcc tcc ttc ccg       407
Tyr Thr Asp Met Tyr Cys Pro Ile Cys Leu His Gln Ala Ser Phe Pro
    40                  45                  50
gtg gag acc aac tgt gga cat ctt ttt tgt ggt gcc tgc att att gct       455
Val Glu Thr Asn Cys Gly His Leu Phe Cys Gly Ala Cys Ile Ile Ala
55                  60                  65                  70
tac tgg cga tat ggt tca tgg ctt ggg gca atc agt tgt cca atc tgt       503
Tyr Trp Arg Tyr Gly Ser Trp Leu Gly Ala Ile Ser Cys Pro Ile Cys
                75                  80                  85
aga caa acg aga cat ggc cac att gca ttg tcc aga aca gct               545
Arg Gln Thr Arg His Gly His Ile Ala Leu Ser Arg Thr Ala
            90                  95                  100
tagaccatga cagttagcat cgaagccacc tgaggaggga ggcagtaacc ttactcctaa    605
cagtatttgg tgaagatgat cagtctcagg atgttctgag attgcatcag gatattaatg    665
attataaccg gagattctca gggcaaccca gatctgtaag taatgctaaa gcatgttcaa    725
```

agttagagga agacacattt cttctctttt gtaaagtgag gtttaccaac aagtattctt 785
tgactatgag aaatcttggc caggcacagt agctaacgcc tataatccta gcactttggg 845
aggccaaggc aggtggatca cttgagccca ggagtttgag accagccttg gaaacatgat 905
gaaaccccat ctctagaaaa aacaccaaaa aattggacaa gagtgttggc acatgcctgt 965
agtccctgct tcttgggagg ctgaaatggg aggatcacct gagcccagga ggttgaggct 1025
atagtgagcc atgatcgcac tattgcactc ccacctgggt ggcagtgaga cccttcctca 1085
aaaaacaaga aaagaaaaaa aaaaaaaaaa a 1116

<210> 32
<211> 1114
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 117..362

<400> 32
ataaggggac gtctagtggg ttgcccggga ggggtggcgg gagcggtcct ggaaataatc 60
tgtcctctgt cgccgggaac tggcgaggta gttccttcgc ggtggagaga cctgga atg 119
Met
1
gcc aaa tat caa ggt gaa gtt caa agt ttg aaa ctg gat gat gat tca 167
Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp Ser
5 10 15
gtt ata gaa gga gta agc gac caa gta ctt gtg gca gtt gtg gtc agt 215
Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val Ser
20 25 30
ttc gct ttg att gct acc ctg gta tat gca ctt ttc aga aat gta cat 263
Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val His
35 40 45
caa aac att cac cca gaa aac cag gag cta gta agg gta ctt cga gaa 311
Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg Glu
50 55 60 65
cag ctt caa aca gaa cag gat gca cct gct gac tcg aca gca gtt cta 359
Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Asp Ser Thr Ala Val Leu
70 75 80
cac tgacatgtac tgtcccatct gcctgcacca agcctccttc ccggtggaga 412
His
ccaactgtgg acatcttttt tgtggtgcct gcattattgc ttactggcga tatggttcat 472
ggcttggggc aatcagttgt ccaatctgta gacaaacgag acatggccac attgcattgt 532
ccagaacagc ttagaccatg acagttagca tcgaagccac ctgaggaggg aggcagtaac 592
cttactccta acagtatttg gtgaagatga tcagtctcag gatgttctga gattgcatca 652
ggatattaat gattataacc ggagattctc agggcaaccc agatctgtaa gtaatgctaa 712
agcatgttca aagttagagg aagacacatt tcttctcttt tgtaaagtga ggtttaccaa 772
caagtattct ttgactatga gaaatcttgg ccaggcacag tagctaacgc ctataatcct 832
agcactttgg gaggccaagg caggtggatc acttgagccc aggagtttga gaccagcctt 892
ggaaacatga tgaaacccca tctctagaaa aaacaccaaa aaattggaca agagtgttgg 952
cacatgcctg tagtccctgc ttcttgggag gctgaaatgg gaggatcacc tgagcccagg 1012
aggttgaggc tatagtgagc catgatcgca ctattgcact cccacctggg tggcagtgag 1072
acccttcctc aaaaaacaag aaaagaaaaa aaaaaaaaaa aa 1114

<210> 33
<211> 2072
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 144..1262

<220>
<221> sig_peptide

<222> 144..224
<223> Von Heijne matrix
score 5.14258625256317
seq FLCQRLVLSTLSG/RP

<400> 33

```
acgtggacgc gtctgggctg ctggaggcag cccgagccgc cgccgtcggt gtcgccgcca      60
ccaccaccat cggagtcacg agtcccgcgt ctgtccgaag tcgccgctct cgggctgctc     120
acgtctcttc ggagagcgcg cac atg gcg act cag gcg cac tcc ctc agc tac     173
                         Met Ala Thr Gln Ala His Ser Leu Ser Tyr
                                    -25                 -20
gca ggg tgc aac ttc ttg tgc caa cgt ctg gtc ctg tct acc ctg agc      221
Ala Gly Cys Asn Phe Leu Cys Gln Arg Leu Val Leu Ser Thr Leu Ser
        -15                 -10                 -5
ggg cgc ccc gtc aaa atc cga aag att cgg gcc aga gac gac aac ccg      269
Gly Arg Pro Val Lys Ile Arg Lys Ile Arg Ala Arg Asp Asp Asn Pro
    1               5                  10                  15
ggc ctc cga gat ttt gaa gcc agc ttc ata agg cta ttg gac aaa ata      317
Gly Leu Arg Asp Phe Glu Ala Ser Phe Ile Arg Leu Leu Asp Lys Ile
                20                  25                  30
acg aat ggt tct cga att gaa ata aac caa aca gga aca acc tta tat      365
Thr Asn Gly Ser Arg Ile Glu Ile Asn Gln Thr Gly Thr Thr Leu Tyr
            35                  40                  45
tat cag cct ggc ctc ctg tat ggt gga tct gtg gaa cat gac tgt agc      413
Tyr Gln Pro Gly Leu Leu Tyr Gly Gly Ser Val Glu His Asp Cys Ser
        50                  55                  60
gtc ctt cgt ggc att ggg tat tac ctg gag agt ctt ctt tgc ttg gct      461
Val Leu Arg Gly Ile Gly Tyr Tyr Leu Glu Ser Leu Leu Cys Leu Ala
    65                  70                  75
cca ttt atg aag cac ccg tta aaa ata gtt cta cga gga gtg acc aat      509
Pro Phe Met Lys His Pro Leu Lys Ile Val Leu Arg Gly Val Thr Asn
80                  85                  90                  95
gat cag att gac cct tca gtt gat gtt ctt aag gca aca gca ctc cct      557
Asp Gln Ile Asp Pro Ser Val Asp Val Leu Lys Ala Thr Ala Leu Pro
                100                 105                 110
ttg ttg aaa caa ttt ggg att gat ggt gaa tca ttt gaa ctg aag att      605
Leu Leu Lys Gln Phe Gly Ile Asp Gly Glu Ser Phe Glu Leu Lys Ile
            115                 120                 125
gtg cga cgg gga atg cct ccc gga gga gga ggc gaa gtg gtt ttc tca      653
Val Arg Arg Gly Met Pro Pro Gly Gly Gly Gly Glu Val Val Phe Ser
        130                 135                 140
tgt cct gtg agg aag gtc ttg aag ccc att caa ctc aca gat cca gga      701
Cys Pro Val Arg Lys Val Leu Lys Pro Ile Gln Leu Thr Asp Pro Gly
    145                 150                 155
aaa atc aaa cgt att aga gga atg gcg tac tct gta cgt gtg tca cct      749
Lys Ile Lys Arg Ile Arg Gly Met Ala Tyr Ser Val Arg Val Ser Pro
160                 165                 170                 175
cag atg gcg aac cgg att gtg gat tct gca agg agc atc ctc aac aag      797
Gln Met Ala Asn Arg Ile Val Asp Ser Ala Arg Ser Ile Leu Asn Lys
                180                 185                 190
ttc ata cct gat atc tat att tac aca gat cac att aaa gga gtc aac      845
Phe Ile Pro Asp Ile Tyr Ile Tyr Thr Asp His Ile Lys Gly Val Asn
            195                 200                 205
tct ggg aag tct ccg ggc ttt ggg ttg tca ctg gtt gct gag acc acc      893
Ser Gly Lys Ser Pro Gly Phe Gly Leu Ser Leu Val Ala Glu Thr Thr
        210                 215                 220
agt ggc acc ttc ctc agt gct gaa ctg gcc tcc aac ccc cag ggc cag      941
Ser Gly Thr Phe Leu Ser Ala Glu Leu Ala Ser Asn Pro Gln Gly Gln
    225                 230                 235
gga gca gca gta ctt cca gag gac ctt ggc agg aac tgt gcc cgg ctg      989
Gly Ala Ala Val Leu Pro Glu Asp Leu Gly Arg Asn Cys Ala Arg Leu
240                 245                 250                 255
ctg ctg gag gaa atc tac agg ggt gga tgc gta gac tcg acc aac caa     1037
```

```
Leu Leu Glu Glu Ile Tyr Arg Gly Gly Cys Val Asp Ser Thr Asn Gln
                260                 265                 270
agc ctg gcg cta cta ctc atg acc ctt gga cag cag gat gtt tcc aaa     1085
Ser Leu Ala Leu Leu Leu Met Thr Leu Gly Gln Gln Asp Val Ser Lys
            275                 280                 285
gtc ctg cta ggc cct ctc tct ccc tac acg ata gaa ttt ttg cgg cat     1133
Val Leu Leu Gly Pro Leu Ser Pro Tyr Thr Ile Glu Phe Leu Arg His
        290                 295                 300
ttg aag agc ttt ttc cag att atg ttt aaa att gaa acc aag cca tgt     1181
Leu Lys Ser Phe Phe Gln Ile Met Phe Lys Ile Glu Thr Lys Pro Cys
    305                 310                 315
ggt gaa gaa ctc aag ggt ggg gat aaa gtg ctg atg acc tgt gtt ggc     1229
Gly Glu Glu Leu Lys Gly Gly Asp Lys Val Leu Met Thr Cys Val Gly
320                 325                 330                 335
att ggt ttc tcc aac ctt agc agg acc ctc aag tgataaccat cacaagataa   1282
Ile Gly Phe Ser Asn Leu Ser Arg Thr Leu Lys
                340                 345
ggccccagtg cctacagaca aagcagaagc tgccacggac accaatggga ccaagtccaa   1342
atggattaat ccaggacaga atagccactt gcttaatttt ctgtgaagaa atatcaatat   1402
acaaataaaa gacatccctg tagcatatgg tttccagctg tttctccagt ggcattgcca   1462
ttgcccagga ggggcccagt caccatgaga gctcccttgc cttacctgga ggaagaatgt   1522
gccttcaggc cacagtcgtg ctgctagaac agtctcgtag ctgcagttca gctgtgcttc   1582
ctcagcctac tatcataggc ttcctcagcc ctctgtcata tggctgtttt gcaaacctgt   1642
ggagtctgtt actgttcttt ctgcaaggac tcacctcctt gagccttggt ttttgttgta   1702
gggattaaat gagataatat gagtggcagc tcttcatgag tcctgcagtg ctaagcaaat   1762
gtcagaaatt ggtgtattag actatttatc tttgatcttc tgaatggatt gctgtcatgg   1822
acacggacac ggatcttcat ctggttcatt gtatttatat gtgagggatg gatggctgcg   1882
gggctccaag taagttattg ggatgttttt atattccagg tgtgctgtac gttcttattt   1942
tattttcaca atagctctgt gatgtaagtg ctatctccat gagaaaattc ataaagggtg   2002
ttttgttcat ttgaaatgta taatgtaaag acattaaatc tcctcattta aggaaaaaaa   2062
aaaaaaaaaa                                                          2072

<210> 34
<211> 409
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 35..316

<220>
<221> sig_peptide
<222> 35..109
<223> Von Heijne matrix
      score 5.38058532480537
      seq AVTSLLSPTPATA/LA

<400> 34
ttttttttcga gaccggaagt gagtgatcga aagc atg gcg tcg gtg gtg ttg gcg     55
                                      Met Ala Ser Val Val Leu Ala
                                      -25                 -20
ctg agg acc cgg aca gcc gtt aca tcc ttg cta agc ccc act ccg gct      103
Leu Arg Thr Arg Thr Ala Val Thr Ser Leu Leu Ser Pro Thr Pro Ala
            -15                 -10                 -5
aca gct ctt gct gtc aga tac gca tcc aag aag tcg ggt ggt agc tcc      151
Thr Ala Leu Ala Val Arg Tyr Ala Ser Lys Lys Ser Gly Gly Ser Ser
        1               5                   10
aaa aac ctc ggt gga aag tca tca ggc aga cgc caa ggc att aag aaa      199
Lys Asn Leu Gly Gly Lys Ser Ser Gly Arg Arg Gln Gly Ile Lys Lys
15                  20                  25                  30
atg gaa ggt cac tat gtt cat gct ggg aac atc att gca aca cag cgc      247
Met Glu Gly His Tyr Val His Ala Gly Asn Ile Ile Ala Thr Gln Arg
```

```
                35                  40                  45
cat ttc cgc tgg cac cca ggt gcc cat gtg agt tgc tcc gtt gct gcc      295
His Phe Arg Trp His Pro Gly Ala His Val Ser Cys Ser Val Ala Ala
            50                  55                  60
ccc ctt ttt cct ttt cta ggt tgacctctcc ttgcccctaa gcatggtaat        346
Pro Leu Phe Pro Phe Leu Gly
        65
aacagttgca tgtattgagt gcttaccaaa tggcaagcat tgtgccaaaa aaaaaaaaaa    406
aaa                                                                  409

<210> 35
<211> 836
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 177..767

<220>
<221> sig_peptide
<222> 177..236
<223> Von Heijne matrix
      score 6.51720597568932
      seq LAVILTLLGLAIL/AI

<400> 35
aatctgctcc acgcaatttc tcagtgatcc tctgcatctc tgcctacaag ggcctccctg     60
acacccaagt tcatattgct cagaaacagt gaacttgagt ttttcatttt accttgatct    120
ctctctgaca aagaaatcca gatgatgcga gacctgatga agacaataca tggaaa atg    179
                                                              Met
                                                              -20
aca gtc ttg gaa ata act ttg gct gtc atc ctg act cta ctg gga ctt      227
Thr Val Leu Glu Ile Thr Leu Ala Val Ile Leu Thr Leu Leu Gly Leu
            -15                 -10                 -5
gcc atc ctg gct att ttg tta aca aga tgg gca cga cgt aag caa agt      275
Ala Ile Leu Ala Ile Leu Leu Thr Arg Trp Ala Arg Arg Lys Gln Ser
            1               5                   10
gaa atg tat atc tcc aga tac agt tca gaa caa agt gct aga ctt ctg      323
Glu Met Tyr Ile Ser Arg Tyr Ser Ser Glu Gln Ser Ala Arg Leu Leu
    15                  20                  25
gac tat gag gat ggt aga gga tcc cga cat gca tat tca aca caa agt      371
Asp Tyr Glu Asp Gly Arg Gly Ser Arg His Ala Tyr Ser Thr Gln Ser
30                  35                  40                  45
gag aga tcc aaa aga gat tac aca cca tca acc aac tct cta gca ctg      419
Glu Arg Ser Lys Arg Asp Tyr Thr Pro Ser Thr Asn Ser Leu Ala Leu
                50                  55                  60
tct cga tca agt att gct tta cct caa gga tcc atg agt agt ata aaa      467
Ser Arg Ser Ser Ile Ala Leu Pro Gln Gly Ser Met Ser Ser Ile Lys
            65                  70                  75
tgt tta caa aca act gaa gaa cct cct tcc aga act gca gga gcc atg      515
Cys Leu Gln Thr Thr Glu Glu Pro Pro Ser Arg Thr Ala Gly Ala Met
        80                  85                  90
atg caa ttc aca gcc cct att ccc gga gct aca gga cct atc aag ctc      563
Met Gln Phe Thr Ala Pro Ile Pro Gly Ala Thr Gly Pro Ile Lys Leu
    95                  100                 105
tct caa aaa acc att gtg caa act cta gga cct att gta caa tat cct      611
Ser Gln Lys Thr Ile Val Gln Thr Leu Gly Pro Ile Val Gln Tyr Pro
110                 115                 120                 125
gga tcc aat ggg agg ata aac ata agc cag ctc acc tca gag gat ctc      659
Gly Ser Asn Gly Arg Ile Asn Ile Ser Gln Leu Thr Ser Glu Asp Leu
                130                 135                 140
act ggg gct aaa gga agg gtc aca tct ggt cca cag ttc cct aat agc      707
```

```
Thr Gly Ala Lys Gly Arg Val Thr Ser Gly Pro Gln Phe Pro Asn Ser
            145                 150                 155
cac cat gtg cca gag aat cta cat gga tac atg aat tcc ctt tcc ctt      755
His His Val Pro Glu Asn Leu His Gly Tyr Met Asn Ser Leu Ser Leu
        160                 165                 170
ttc tcc cct gct tgactccctc tcccttatgt gtaaacaatt taaaaatatg          807
Phe Ser Pro Ala
    175
atagtgtata aatgaaaaaa aaaaaaaaa                                     836

<210> 36
<211> 1323
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 208..1239

<220>
<221> sig_peptide
<222> 208..294
<223> Von Heijne matrix
      score 5.73027134157378
      seq GLVLICVCSKTHS/LK

<400> 36
agtctcgtat cgcgcccggg aggcgccgga gcccagcggc tggcgccaga tccaggctcc     60
tggaagaacc atgtccggca gctactggtc atgccaggca cacactgctg cccaagagga    120
gctgctgttt gaattatctg tgaatgttgg gaagaggaat gccagagctg ccggctgaaa    180
attacccaac caagagaaat ctgcagg atg gac ttt ctg gtc ctc ttc ttg ttc    234
                              Met Asp Phe Leu Val Leu Phe Leu Phe
                                              -25
tac ctg gct tcg gtg ctg atg ggt ctt gtt ctt atc tgc gtc tgc tcg      282
Tyr Leu Ala Ser Val Leu Met Gly Leu Val Leu Ile Cys Val Cys Ser
-20                 -15                 -10                 -5
aaa acc cat agc ttg aaa ggc ctg gcc agg gga gga gca cag ata ttt      330
Lys Thr His Ser Leu Lys Gly Leu Ala Arg Gly Gly Ala Gln Ile Phe
                1                   5                   10
tcc tgt ata att cca gaa tgt ctt cag aga gcc gtg cat gga ttg ctt      378
Ser Cys Ile Ile Pro Glu Cys Leu Gln Arg Ala Val His Gly Leu Leu
        15                  20                  25
cat tac ctt ttc cat acg aga aac cac acc ttc att gtc ctg cac ctg      426
His Tyr Leu Phe His Thr Arg Asn His Thr Phe Ile Val Leu His Leu
    30                  35                  40
gtc ttg caa ggg atg gtt tat act gag tac acc tgg gaa gta ttt ggc      474
Val Leu Gln Gly Met Val Tyr Thr Glu Tyr Thr Trp Glu Val Phe Gly
45                  50                  55                  60
tac tgt cag gag ctg gag ttg tcc ttg cat tac ctt ctt ctg ccc tat      522
Tyr Cys Gln Glu Leu Glu Leu Ser Leu His Tyr Leu Leu Leu Pro Tyr
                65                  70                  75
ctg ctg cta ggt gta aac ctg ttt ttt ttc acc ctg act tgt gga acc      570
Leu Leu Leu Gly Val Asn Leu Phe Phe Phe Thr Leu Thr Cys Gly Thr
            80                  85                  90
aat cct ggc att ata aca aaa gca aat gaa tta tta ttt ctt cat gtt      618
Asn Pro Gly Ile Ile Thr Lys Ala Asn Glu Leu Leu Phe Leu His Val
        95                  100                 105
tat gaa ttt gat gaa gtg atg ttt cca aag aac gtg agg tgc tct act      666
Tyr Glu Phe Asp Glu Val Met Phe Pro Lys Asn Val Arg Cys Ser Thr
    110                 115                 120
tgt gat tta agg aaa cca gct cga tcc aag cac tgc agt gtg tgt aac      714
Cys Asp Leu Arg Lys Pro Ala Arg Ser Lys His Cys Ser Val Cys Asn
125                 130                 135                 140
```

```
tgg tgt gtg cac cgt ttc gac cat cac tgt gtt tgg gtg aac aac tgc      762
Trp Cys Val His Arg Phe Asp His His Cys Val Trp Val Asn Asn Cys
                145                 150                 155
atc ggg gcc tgg aac atc agg tac ttc ctc atc tac gtc ttg acc ttg      810
Ile Gly Ala Trp Asn Ile Arg Tyr Phe Leu Ile Tyr Val Leu Thr Leu
            160                 165                 170
acg gcc tcg gct gcc acc gtc gcc att gtg agc acc act ttt ctg gtc      858
Thr Ala Ser Ala Ala Thr Val Ala Ile Val Ser Thr Thr Phe Leu Val
        175                 180                 185
cac ttg gtg gtg atg tca gat tta tac cag gag act tac atc gat gac      906
His Leu Val Val Met Ser Asp Leu Tyr Gln Glu Thr Tyr Ile Asp Asp
    190                 195                 200
ctt gga cac ctc cat gtt atg gac acg gtc ttt ctt att cag tac ctg      954
Leu Gly His Leu His Val Met Asp Thr Val Phe Leu Ile Gln Tyr Leu
205                 210                 215                 220
ttc ctg act ttt cca cgg att gtc ttc atg ctg ggc ttt gtc gtg gtt     1002
Phe Leu Thr Phe Pro Arg Ile Val Phe Met Leu Gly Phe Val Val Val
                225                 230                 235
ctg agc ttc ctc ctg ggt ggc tac ctg ttg ttt gtc ctg tat ctg gcg     1050
Leu Ser Phe Leu Leu Gly Gly Tyr Leu Leu Phe Val Leu Tyr Leu Ala
            240                 245                 250
gcc acc aac cag act act aac gag tgg tac aga ggt gac tgg gcc tgg     1098
Ala Thr Asn Gln Thr Thr Asn Glu Trp Tyr Arg Gly Asp Trp Ala Trp
        255                 260                 265
tgc cag cgt tgt ccc ctt gtg gcc tgg cct ccg tca gca gag ccc caa     1146
Cys Gln Arg Cys Pro Leu Val Ala Trp Pro Pro Ser Ala Glu Pro Gln
    270                 275                 280
gtc cac cgg aac att cac tcc cat ggg ctt cgg agc aac ctt caa gag     1194
Val His Arg Asn Ile His Ser His Gly Leu Arg Ser Asn Leu Gln Glu
285                 290                 295                 300
atc ttt cta cct gcc ttt cca tgt cat gag agg aag aaa caa gaa          1239
Ile Phe Leu Pro Ala Phe Pro Cys His Glu Arg Lys Lys Gln Glu
                305                 310                 315
tgacaagtgt atgactgcct ttgagctgta gttcccgttt atttacacat gtggatcctc    1299
gttttccaaa aaaaaaaaaa aaaa                                          1323

<210> 37
<211> 1945
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 60..1682

<220>
<221> sig_peptide
<222> 60..143
<223> Von Heijne matrix
      score 3.75144398608723
      seq SGLLLQVLFRLIT/FV

<400> 37
atcgcgacta aacggagtgg cggcggcatt tcctggtgtc tgagcctggc gcggaggct      59
atg ggc agc cag gag gtg ctg ggc cac gcg gcc cgg ctg tcc tcc tcc      107
Met Gly Ser Gln Glu Val Leu Gly His Ala Ala Arg Leu Ser Ser Ser
            -25                 -20                 -15
ggt ctc ctc ctg cag gtg ttg ttt cgg ttg atc acc ttt gtc ttg aat      155
Gly Leu Leu Leu Gln Val Leu Phe Arg Leu Ile Thr Phe Val Leu Asn
        -10                 -5                  1
gca ttt att ctt cgc ttc ctg tca aag gaa atc gtt ggc gta gta aat      203
Ala Phe Ile Leu Arg Phe Leu Ser Lys Glu Ile Val Gly Val Val Asn
5                   10                  15                  20
```

```
gta aga cta acg ctg ctt tac tca acc acc ctc ttc ctg gcc aga gag      251
Val Arg Leu Thr Leu Leu Tyr Ser Thr Thr Leu Phe Leu Ala Arg Glu
                25                  30                  35
gcc ttc cgc aga gca tgt ctc agt ggg ggc acc cag cga gac tgg agc      299
Ala Phe Arg Arg Ala Cys Leu Ser Gly Gly Thr Gln Arg Asp Trp Ser
            40                  45                  50
cag acc ctc aac ctg ctg tgg cta aca gtc ccc ctg ggt gtg ttt tgg      347
Gln Thr Leu Asn Leu Leu Trp Leu Thr Val Pro Leu Gly Val Phe Trp
        55                  60                  65
tcc tta ttc ctg ggc tgg atc tgg ttg cag ctg ctt gaa gtg cct gat      395
Ser Leu Phe Leu Gly Trp Ile Trp Leu Gln Leu Leu Glu Val Pro Asp
    70                  75                  80
cct aat gtt gtc cct cac tat gca act gga gtg gtg ctg ttt ggt ctc      443
Pro Asn Val Val Pro His Tyr Ala Thr Gly Val Val Leu Phe Gly Leu
85                  90                  95                  100
tcg gca gtg gtg gag ctt cta gga gag ccc ttt tgg gtc ttg gca caa      491
Ser Ala Val Val Glu Leu Leu Gly Glu Pro Phe Trp Val Leu Ala Gln
                105                 110                 115
gca cat atg ttt gtg aag ctc aag gtg att gca gag agc ctg tcg gta      539
Ala His Met Phe Val Lys Leu Lys Val Ile Ala Glu Ser Leu Ser Val
            120                 125                 130
att ctt aag acc gtt ctg aca gct ttt ctc gtg ctg tgg ttg cct cac      587
Ile Leu Lys Thr Val Leu Thr Ala Phe Leu Val Leu Trp Leu Pro His
        135                 140                 145
tgg gga ttg tac att ttc tct ttg gcc cag ctt ttc tat acc aca gtt      635
Trp Gly Leu Tyr Ile Phe Ser Leu Ala Gln Leu Phe Tyr Thr Thr Val
    150                 155                 160
ctg gtg ctc tgc tat gtt att tat ttc aca aag tta ctg ggt tcc cca      683
Leu Val Leu Cys Tyr Val Ile Tyr Phe Thr Lys Leu Leu Gly Ser Pro
165                 170                 175                 180
gaa tca acc aag ctt caa act ctt cct gtc tcc aga ata aca gat ctg      731
Glu Ser Thr Lys Leu Gln Thr Leu Pro Val Ser Arg Ile Thr Asp Leu
                185                 190                 195
tta ccc aat att aca aga aat gga gcg ttt ata aac tgg aaa gag gct      779
Leu Pro Asn Ile Thr Arg Asn Gly Ala Phe Ile Asn Trp Lys Glu Ala
            200                 205                 210
aaa ctg act tgg agt ttt ttc aaa cag tct ttc ttg aaa cag att ttg      827
Lys Leu Thr Trp Ser Phe Phe Lys Gln Ser Phe Leu Lys Gln Ile Leu
        215                 220                 225
aca gaa ggc gag cga tat gtg atg aca ttt ttg aat gta ttg aac ttt      875
Thr Glu Gly Glu Arg Tyr Val Met Thr Phe Leu Asn Val Leu Asn Phe
    230                 235                 240
ggt gat cag ggt gtg tat gat ata gtg aat aat ctt ggc tcc ctt gtg      923
Gly Asp Gln Gly Val Tyr Asp Ile Val Asn Asn Leu Gly Ser Leu Val
245                 250                 255                 260
gcc aga tta att ttc cag cca ata gag gaa agt ttt tat ata ttt ttt      971
Ala Arg Leu Ile Phe Gln Pro Ile Glu Glu Ser Phe Tyr Ile Phe Phe
                265                 270                 275
gct aag gtg ctg gag agg gga aag gat gcc aca ctt cag aag cag gag     1019
Ala Lys Val Leu Glu Arg Gly Lys Asp Ala Thr Leu Gln Lys Gln Glu
            280                 285                 290
gac gtt gct gtg gct gct gca gtc ttg gag tcc ctg ctc aag ctg gcc     1067
Asp Val Ala Val Ala Ala Ala Val Leu Glu Ser Leu Leu Lys Leu Ala
        295                 300                 305
ctg ctg gcc ggc ctg acc atc act gtt ttt ggc ttt gcc tat tct cag     1115
Leu Leu Ala Gly Leu Thr Ile Thr Val Phe Gly Phe Ala Tyr Ser Gln
    310                 315                 320
ctg gct ctg gat atc tac gga ggg acc atg ctt agc tca gga tcc ggt     1163
Leu Ala Leu Asp Ile Tyr Gly Gly Thr Met Leu Ser Ser Gly Ser Gly
325                 330                 335                 340
cct gtt ttg ctg cgt tcc tac tgt ctc tat gtt ctc ctg ctt gcc atc     1211
Pro Val Leu Leu Arg Ser Tyr Cys Leu Tyr Val Leu Leu Leu Ala Ile
                345                 350                 355
```

```
aat gga gtg aca gag tgt ttc aca ttt gct gcc atg agc aaa gag gag     1259
Asn Gly Val Thr Glu Cys Phe Thr Phe Ala Ala Met Ser Lys Glu Glu
            360                 365                 370
gtc gac agg tac aat ttt gtg atg ctg gcc ctg tcc tcc tca ttc ctg     1307
Val Asp Arg Tyr Asn Phe Val Met Leu Ala Leu Ser Ser Ser Phe Leu
        375                 380                 385
gtg tta tcc tat ctc ttg acc cgt tgg tgt ggc agc gtg ggc ttc atc     1355
Val Leu Ser Tyr Leu Leu Thr Arg Trp Cys Gly Ser Val Gly Phe Ile
    390                 395                 400
ttg gcc aac tgc ttt aac atg ggc att cgg atc acg cag agc ctt tgc     1403
Leu Ala Asn Cys Phe Asn Met Gly Ile Arg Ile Thr Gln Ser Leu Cys
405                 410                 415                 420
ttc atc cac cgc tac tac cga agg agc ccc cac agg ccc ctg gct ggc     1451
Phe Ile His Arg Tyr Tyr Arg Arg Ser Pro His Arg Pro Leu Ala Gly
                425                 430                 435
ctg cac cta tcg cca gtc ctg ctc ggg aca ttt gcc ctc agt ggt ggg     1499
Leu His Leu Ser Pro Val Leu Leu Gly Thr Phe Ala Leu Ser Gly Gly
            440                 445                 450
gtt act gct gtt tcg gag gta ttc ctc tgc tgt gag cag ggc tgg cca     1547
Val Thr Ala Val Ser Glu Val Phe Leu Cys Cys Glu Gln Gly Trp Pro
        455                 460                 465
gcc aga ctg gca cac att gct gtg ggg gcc ttc tgt ctg gga gca act     1595
Ala Arg Leu Ala His Ile Ala Val Gly Ala Phe Cys Leu Gly Ala Thr
    470                 475                 480
ctc ggg aca gca ttc ctc aca gag acc aag ctg atc cat ttc ctc agg     1643
Leu Gly Thr Ala Phe Leu Thr Glu Thr Lys Leu Ile His Phe Leu Arg
485                 490                 495                 500
act cag tta ggt gtg ccc aga cgc act gac aaa atg aca tgacttcagg      1692
Thr Gln Leu Gly Val Pro Arg Arg Thr Asp Lys Met Thr
                505                 510
gaagcctgga cacccgaggc acctggacca gctatgggta gttctgtggg tggaacacat   1752
tctgtgtaag agccccactg agggctctgc agcggagtga cagcaacccc agagatgagg   1812
caccagagag tgccactgca tgagacacct gtgaccattc gaagtctgaa atgcgggggg   1872
ggagtttcat ttttaagtga agaccaaaag ccctttaaaa ataatagttt tttatcaaaa   1932
aaaaaaaaaa aaa                                                      1945

<210> 38
<211> 1330
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 198..998

<220>
<221> sig_peptide
<222> 198..269
<223> Von Heijne matrix
      score 9.08017839002281
      seq LLLGPGLLATVRA/EC

<400> 38
agaaatcagc cctttgcaga gggcgcagag ggcctggaaa cctctgggac cttttcccag     60
gaactgttta tggtttcccc ctaggtctag gagacgtaga tgcataggtg gattggatac    120
atcgatggta gctataagag tcgtgtctga acccggcttt tccaattggc ctgctccatc    180
cgaacagcgt caactcc atg gcg cgg ttc ctg aca ctt tgc act tgg ctg       230
                  Met Ala Arg Phe Leu Thr Leu Cys Thr Trp Leu
                                  -20                 -15
ctg ttg ctc ggc ccc ggg ctc ctg gcg acc gtg cgg gcc gaa tgc agc      278
Leu Leu Leu Gly Pro Gly Leu Leu Ala Thr Val Arg Ala Glu Cys Ser
            -10                 -5                  1
cag gat tgc gcg acg tgc agc tac cgc cta gtg cgc ccg gcc gac atc      326
```

```
Gln Asp Cys Ala Thr Cys Ser Tyr Arg Leu Val Arg Pro Ala Asp Ile
    5                  10                  15
aac ttc ctg gct tgc gta atg gaa tgt gaa ggt aaa ctg cct tct ctg      374
Asn Phe Leu Ala Cys Val Met Glu Cys Glu Gly Lys Leu Pro Ser Leu
20                  25                  30                  35
aaa att tgg gaa acc tgc aag gag ctc ctg cag ctg tcc aaa cca gat      422
Lys Ile Trp Glu Thr Cys Lys Glu Leu Leu Gln Leu Ser Lys Pro Asp
                40                  45                  50
ctt cct caa gat ggc acc agc acc ctc aga gaa aat agc aaa ccg gaa      470
Leu Pro Gln Asp Gly Thr Ser Thr Leu Arg Glu Asn Ser Lys Pro Glu
            55                  60                  65
gaa agc cat ttg cta gcc aaa agg tat ggg ggc ttc atg aaa agg tat      518
Glu Ser His Leu Leu Ala Lys Arg Tyr Gly Gly Phe Met Lys Arg Tyr
        70                  75                  80
gga ggc ttc atg aag aaa atg gat gag ctt tat ccc atg gag cca gaa      566
Gly Gly Phe Met Lys Lys Met Asp Glu Leu Tyr Pro Met Glu Pro Glu
    85                  90                  95
gaa gag gcc aat gga agt gag atc ctc gcc aag cgg tat ggg ggc ttc      614
Glu Glu Ala Asn Gly Ser Glu Ile Leu Ala Lys Arg Tyr Gly Gly Phe
100                 105                 110                 115
atg aag aag gat gca gag gag gac gac tcg ctg gcc aat tcc tca gac      662
Met Lys Lys Asp Ala Glu Glu Asp Asp Ser Leu Ala Asn Ser Ser Asp
                120                 125                 130
ctg cta aaa gag ctt ctg gaa aca ggg gac aac cga gag cgt agc cac      710
Leu Leu Lys Glu Leu Leu Glu Thr Gly Asp Asn Arg Glu Arg Ser His
            135                 140                 145
cac cag gat ggc agt gat aat gag gaa gaa gtg agc aag aga tat ggg      758
His Gln Asp Gly Ser Asp Asn Glu Glu Glu Val Ser Lys Arg Tyr Gly
        150                 155                 160
ggc ttc atg aga ggc tta aag aga agc ccc caa ctg gaa gat gaa gcc      806
Gly Phe Met Arg Gly Leu Lys Arg Ser Pro Gln Leu Glu Asp Glu Ala
    165                 170                 175
aaa gag ctg cag aag cga tat ggg ggc ttc atg aga aga gta ggt cgc      854
Lys Glu Leu Gln Lys Arg Tyr Gly Gly Phe Met Arg Arg Val Gly Arg
180                 185                 190                 195
cca gag tgg tgg atg gac tac cag aaa cgg tat gga ggt ttc ctg aag      902
Pro Glu Trp Trp Met Asp Tyr Gln Lys Arg Tyr Gly Gly Phe Leu Lys
                200                 205                 210
cgc ttt gcc gag gct ctg ccc tcc gac gaa gaa ggc gaa agt tac tcc      950
Arg Phe Ala Glu Ala Leu Pro Ser Asp Glu Glu Gly Glu Ser Tyr Ser
            215                 220                 225
aaa gaa gtt cct gaa atg gaa aaa aga tac gga gga ttt atg aga ttt      998
Lys Glu Val Pro Glu Met Glu Lys Arg Tyr Gly Gly Phe Met Arg Phe
        230                 235                 240
taatattttt cccactagtg gccccaggcc ccagcaagcc tccctccatc ctccagtggg   1058
aaactgttga tggtgtttta ttgtcatgtg ttgcttgcct tgtatagttg acttcattgt   1118
ctggataact atacaacctg aaaactgtca tttcaggttc tgtgctcttt ttggagtctt   1178
taagctcagt attagtctat tgcagctatc tcgttttcat gctaaaatag tttttgttat   1238
cttgtctctt atttttgaca aacatcaata aatgcttact tgtatataga gataataaac   1298
ctattacccc aagtgcaaaa aaaaaaaaaa aa                                 1330

<210> 39
<211> 2124
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 505..1590

<220>
<221> sig_peptide
<222> 505..624
```

<223> Von Heijne matrix
score 8.5056444915604
seq VVMLMLLTLLVLG/MV

<400> 39

```
cctggcataa ctgataggca tgtatgggag gaccacattc ctggggacag cctgggtatg      60
tgacatggca ggtgaccagg ttcccatgaa tgcccgaggc tgtgcccatc ccatgagctg     120
gggcttccct ggaggtaaag agctagggtg gggtggcagt gggtagaacc ccagctggac     180
agctccttcc ttagctctgt gattgctaca gctggttctg gaagccacag gcgccctcag     240
gacaaatggg gcttcttcag cacagggtag tgagtgctga gctaagcaag gacactgtcc     300
ccttctctgc ccaggctcga gctgtgcacc tttaccctgg caattgccct gggtgctgtc     360
ctgctcctgc ccttctccat catcagcaat gaggtgctgc tctccctgcc tcggaactac     420
tacatccagt ggctcaacgg ctcctcatc catggcctct ggaaccttgt ttttctcttc     480
tccaacctgt ccctcatctt cctc atg ccc ttt gca tat ttc ttc act gag         531
                          Met Pro Phe Ala Tyr Phe Phe Thr Glu
                          -40                 -35
tct gag ggc ttt gct ggc tcc aga aag ggt gtc ctg ggc cgg gtc tat       579
Ser Glu Gly Phe Ala Gly Ser Arg Lys Gly Val Leu Gly Arg Val Tyr
    -30                 -25                 -20
gag aca gtg gtg atg ttg atg ctc ctc act ctg ctg gtg cta ggt atg       627
Glu Thr Val Val Met Leu Met Leu Leu Thr Leu Leu Val Leu Gly Met
-15                 -10                 -5                  1
gtg tgg gtg gca tca gcc att gtg gac aag aac aag gcc aac aga gag       675
Val Trp Val Ala Ser Ala Ile Val Asp Lys Asn Lys Ala Asn Arg Glu
            5                   10                  15
tca ctc tat gac ttt tgg gag tac tat ctc ccc tac ctc tac tca tgc       723
Ser Leu Tyr Asp Phe Trp Glu Tyr Tyr Leu Pro Tyr Leu Tyr Ser Cys
        20                  25                  30
atc tcc ttc ctt ggg gtt ctg ctg ctc ctg gtg tgt act cca ctg ggt       771
Ile Ser Phe Leu Gly Val Leu Leu Leu Leu Val Cys Thr Pro Leu Gly
    35                  40                  45
ctc gcc cgc atg ttc tcc gtc act ggg aag ctg cta gtc aag ccc cgg       819
Leu Ala Arg Met Phe Ser Val Thr Gly Lys Leu Leu Val Lys Pro Arg
50                  55                  60                  65
ctg ctg gaa gac ctg gag gag cag ctg tac tgc tca gcc ttt gag gag       867
Leu Leu Glu Asp Leu Glu Glu Gln Leu Tyr Cys Ser Ala Phe Glu Glu
                70                  75                  80
gca gcc ctg acc cgc agg atc tgt aat cct act tcc tgc tgg ctg cct       915
Ala Ala Leu Thr Arg Arg Ile Cys Asn Pro Thr Ser Cys Trp Leu Pro
            85                  90                  95
tta gac atg gag ctg cta cac aga cag gtc ctg gct ctg cag aca cag       963
Leu Asp Met Glu Leu Leu His Arg Gln Val Leu Ala Leu Gln Thr Gln
        100                 105                 110
agg gtc ctg ctg gag aag agg cgg aag gct tca gcc tgg caa cgg aac      1011
Arg Val Leu Leu Glu Lys Arg Arg Lys Ala Ser Ala Trp Gln Arg Asn
    115                 120                 125
ctg ggc tac ccc ctg gct atg ctg tgc ttg ctg gtg ctg acg ggc ctg      1059
Leu Gly Tyr Pro Leu Ala Met Leu Cys Leu Leu Val Leu Thr Gly Leu
130                 135                 140                 145
tct gtg ctc att gtg gcc atc cac atc ctg gag ctg ctc atc gat gag      1107
Ser Val Leu Ile Val Ala Ile His Ile Leu Glu Leu Leu Ile Asp Glu
                150                 155                 160
gct gcc atg ccc cga ggc atg cag ggt acc tcc tta ggc cag gtc tcc      1155
Ala Ala Met Pro Arg Gly Met Gln Gly Thr Ser Leu Gly Gln Val Ser
            165                 170                 175
ttc tcc aag ctg ggc tcc ttt ggt gcc gtc att cag gtt gta ctc atc      1203
Phe Ser Lys Leu Gly Ser Phe Gly Ala Val Ile Gln Val Val Leu Ile
        180                 185                 190
ttt tac cta atg gtg tcc tca gtt gtg ggc ttc tat agc tct cca ctc      1251
Phe Tyr Leu Met Val Ser Ser Val Val Gly Phe Tyr Ser Ser Pro Leu
    195                 200                 205
ttc cgg agc ctg cgg ccc aga tgg cac gac act gcc atg acg cag ata      1299
Phe Arg Ser Leu Arg Pro Arg Trp His Asp Thr Ala Met Thr Gln Ile
```